Міністерство освіти і науки, молоді та спорту України
Волинський національний університет імені Лесі Українки

Кормош Ж. О., Супрунович С. В.

Наукова інформація з хімії, її пошук та обробка

Навчальний посібник

Луцьк
Волинський національний університет імені Лесі Українки
2011

УДК 54:004.382.7(075.8)
ББК 24с.я73
К 66

Рекомендовано Міністерством освіти і науки, молоді та спорту України як навчальний посібник для студентів вищих навчальних закладів
(№ 1/11-632 від 25.01.2011 р.)

Рецензенти:

Зайцев Володимир Миколайович – доктор хімічних наук, професор, член-кореспондент НАН України, завідувач кафедри аналітичної хімії Київського національного університету імені Тараса Шевченка;

Логінова Лідія Павлівна – доктор хімічних наук, професор, завідувач кафедри хімічної метрології Харківського національного університету імені В. Каразіна;

Базель Ярослав Рудольфович – доктор хімічних наук, професор, завідувач кафедри аналітичної хімії Ужгородського національного університету

Кормош Ж. О., Супрунович С. В.

К 66 Наукова інформація з хімії, її пошук та обробка: навч. посіб. для студ. вищ. навч. закл. – Луцьк : ВНУ імені Лесі Українки, 2011. – 259 с.

У посібнику розглянуто особливості наукової інформації, її розвиток та способи доступу до неї. Основна увага звернено на специфіку пошуку наукової інформації з хімії в мережі Інтернет.

Для студентів, аспірантів і науковців природничих спеціальностей.

УДК 54:004.738.52
УДК 54:004.382.7(075.8)

ЗМІСТ

ВСТУП .....
2. ПОНЯТТЯ НАУКОВОЇ ІНФОРМАЦІЇ .....
2.1. Властивості наукової інформації .....
2.2. Старіння інформації .....
2.3. Питання для самоперевірки .....
3. СТРУКТУРА ДЖЕРЕЛ НАУКОВОЇ ІНФОРМАЦІЇ .....
4. ДРУКОВАНІ НАУКОВІ ПУБЛІКАЦІЇ .....
4.1. Первинні джерела наукової інформації .....
4.1.1. Наукові журнали .....
4.1.2. Структура наукового журналу .....
4.1.3. Типи публікацій в науковому журналі .....
4.2. Інші джерела наукової інформації .....
4.2.1 Дисертації (Dissertation, Thesis) .....
4.2.2. Наукові звіти (Research Report) .....
4.2.3. Патенти .....
4.2.4. Авторські свідоцтва .....
4.2.5. Відомчі матеріали .....
4.3. Вторинні джерела інформації .....
4.3.1. Сигнальна інформація .....
4.3.2. Анонси .....
4.3.3. Препринти .....
4.3.4. Матеріали конференцій .....
4.3.5. Реферативні журнали .....
4.4. Третинні джерела наукової інформації .....
4.5. Питання для самоперевірки .....
5. СТРУКТУРА ДЖЕРЕЛ ПЕРВИННОЇ НАУКОВОЇ ІНФОРМАЦІЇ .....
5.1. Структура наукової статті .....
5.2. Структура патенту на винахід .....
5.3. Питання для самоперевірки .....
6. БІБЛІОГРАФІЧНИЙ ОПИС .....
6.1. Види бібліографічного опису .....
6.2. Бібліографічний опис електронних документів .....
6.3. Посилання на джерела інформації в тексті .....
6.4. Питання для самоперевірки .....

7. ІСТОРІЯ НАКОПИЧЕННЯ НАУКОВОЇ ІНФОРМАЦІЇ.....
7.1. Зростання кількості публікацій від середини XVII до XX ст.....
7.2. Вплив сонячної активності на роботу вчених .....
7.3. Бібліотечна криза .....
7.4. Комерціалізація інформації з хімії .....
7.5. Питання для самоперевірки .....
8. ОЦІНЮВАННЯ ЕФЕКТИВНОСТІ НАУКОВИХ ПУБЛІКАЦІЙ.....
8.1. Індекс цитувань .....
8.2. Імпакт-фактор.....
8.3. Індекс Хірша.....
8.4. Питання для самоперевірки .....
9. ІНТЕРНЕТ.....
9.1. Історія Інтернету .....
9.2. Протоколи та адреси.....
9.3. Адресація в Інтернеті.....
9.3.1. Структура доменних імен.....
9.3.2. WWW – Word Wide Web .....
9.3.3. Система адрес у мережі Internet.....
9.3.4. Структура веб-адрес.....
9.4. Питання для самоперевірки .....
9.5. Структура Інтернету .....
9.6. Класифікація сервісів Інтернету .....
9.7. Формати файлів з Інтернету.....
9.8. Питання для самоперевірки .....
10. ПОШУКОВІ МАШИНИ.....
10.1. Формування асоціативних запитів .....
10.2. Формулювання асоціативних запитів.....
10.3. Бази даних.....
10.4. Питання для самопідготовки.....
11. ПРИНЦИПИ ФОРМУВАННЯ ПОШУКОВИХ БАЗ ДАНИХ.....
12. ПОШУКОВІ ЗАПИТИ .....
12.1. Мова пошукових запитів .....
12.2. Особливості оформлення пошукових запитів .....
12.3. Питання для самоперевірки .....

13. МЕТОДИ ПОШУКУ ІНФОРМАЦІЇ.....
13.1. Загальні методи пошуку інформації.....
13.2. Питання для самоперевірки .....
14. ПОШУК В ІНТЕРНЕТІ З ДОПОМОГОЮ ПОШУКОВИХ МАШИН.....
14.1. Оцінка якості пошуку .....
14.2. Пошук дублікатів документів.....
14.3. Проблема вибору пошукової машини.....
14.4. Питання для самоперевірки .....
15. РЕАЛІЗАЦІЯ РІЗНИХ МЕТОДІВ ПОШУКУ ІНФОРМАЦІЇ В ІНТЕРНЕТ.....
15.1. Дедуктивний метод.....
15.2. Індуктивний метод.....
15.3. Питання для самоперевірки .....
16. ІНДЕКСАТОРИ І КЛАСИФІКАТОРИ .....
16.1. УДК .....
16.2. ББК .....
16.3. Авторський знак.....
16.4. CAS.....
16.5. PURL.....
16.6. SMILES.....
16.7. Комплексний книготорговий індекс-шифр .....
16.8. ISSN.....
16.9. ISBN –Міжнародний стандартний номер книги .....
16.10.Цифровий ідентифікатор об’єкта (DOI).....
16.11.Міжнародний хімічний ідентифікатор InChI.....
16.12. InChI Key
16.13.Питання для самоперевірки .....
17. ПОШУКОВІ МАШИНИ ЗАГАЛЬНОГО КОРИСТУВАННЯ .....
17.1. Google.....
17.2. Bing.....
17.3. AltaVista .....
17.4. Yahoo!.....
17.5. Yandex .....
17.6. Освітні програми Yandex.....
17.7. Rambler.....
17.8. Питання для самоперевірки .....

18. МЕТАПОШУКОВІ МАШИНИ .....
18.1. Nigma .....
18.1.1. Кластеризація результатів пошуку .....
18.1.2. Розшифрування скорочень .....
18.1.3. Специфічні сервіси й інструменти .....
18.2. Meta .....
18.3. Питання для самоперевірки .....
19. ПОШУКОВІ МАШИНИ НАУКОВОЇ ІНФОРМАЦІЇ .....
19.1. Scirus 209
19.2. Питання для самоперевірки 212
20. ПОШУКОВІ МАШИНИ БІБЛІОГРАФІЧНОЇ НАУКОВОЇ ІНФОРМАЦІЇ .....
20.1. SCOPUS .....
20.2. Питання для самоперевірки .....
21. ІНТЕРНЕТ-ЕНЦИКЛОПЕДІЇ .....
21.1. Вікіпедія .....
21.2. Питання для самоперевірки .....
22. ВЕБ-АРХІВИ .....
22.1. Мультимедійні ресурси .....
22.2. BookServer .....
22.3. Питання для самоперевірки .....
23. ХІМІЧНІ КАТАЛОГИ .....
24. ХІМІЧНІ ПОРТАЛИ .....
25. БАЗИ ДАНИХ ІНФОРМАЦІЇ З ХІМІЇ .....
25.1. PubChem .....
25.2. ChemSpider .....
25.3. NIST .....
26. ПОШУКОВІ МАШИНИ ПАТЕНТНОЇ ІНФОРМАЦІЇ .....
26.1. FIZ Karlsruhe .....
26.2. USPTO .....
26.2.1. Пошук патентів .....
26.2.2. Швидкий пошук .....
26.2.3. Розширений пошук .....
26.3. esp@cenet .....
26.3.1. Простий пошук .....
26.3.2. Розширений пошук .....
26.3.3. Пошук за номером патенту .....

26.3.4 Пошук класифікації.....
26.3.5 Результати пошуку .....
26.3.6 Сторінка документа.....
27. РОБОТА З ДЖЕРЕЛАМИ ІНФОРМАЦІЇ .....

ЛІТЕРАТУРА .....

ВСТУП

Курс “Наукова інформація з хімії, її пошук та обробка” у системі підготовки спеціалістів-хіміків посідає особливе місце. Він є містком між хімією та окремими розділами інформатики, історії, прикладної лінгвістики та інших наук. Вивчення цього курсу сприятиме становленню цілісного уявлення про хімічну науку та формуванню системного підходу під час розгляду процесів поширення наукової інформації у студентів-хіміків.

Пропонований посібник є основною частиною курсу “Наукова інформація з хімії, її пошук та обробка”, котрий упродовж багатьох років читають викладачі хімічного факультету Волинського національного університету імені Лесі Українки.

У виданні значне місце відведено розгляду методів пошуку інформації, які є однією з ланок наукового процесу будь-якої дисципліни. Значну увагу приділено основним засобам пошуку інформації у всесвітній мережі Інтернет. Окремий розділ присвячено історичним аспектам накопичення наукової інформації. Розглянуто принципи формування пошукових баз даних і на їх основі сформовано способи ефективного пошуку інформації. Оволодіння цими методами є важливим етапом у підготовці спеціаліста-хіміка для його успішної діяльності в різних галузях фундаментальної, прикладної хімії та методики її викладання.

Для глибшого засвоєння теоретичного матеріалу в кінці посібника наведено приклади застосування викладених підходів щодо вирішення практичних завдань із пошуку інформації.

Автори висловлюють щиру подяку своїм колегам за цінні поради та рекомендації, які було взято до уваги під час написання цього посібника.

2. ПОНЯТТЯ НАУКОВОЇ ІНФОРМАЦІЇ

Головним результатом діяльності вчених є інформація, яку зазвичай подають у вигляді публікацій.

Єдиного наукового визначення поняття "інформація" немає. На сьогодні існує понад 300 тлумачень цього терміна. Слово "інформація" походить від латинського "informatio", що означає роз'яснення, інформування, зміст повідомлення, відомості включно з їх передаванням у просторі й часі.

Зі змістового погляду "інформація" – це відомості про когось або про щось, а з формального – набір знаків і сигналів.

У кожній з наук слово "інформація" тлумачать по-різному. Один з видів інформації – наукова, у якій своєю чергою виокремлюють науково-технічну.

Наукова інформація (англ. Scientific information, SI) – це логічно організована інформація, яку отримують у процесі наукових досліджень. Вона відображає явища і закони природи, суспільства і мислення.

Науково-технічна інформація (англ. Science and technical information, STI) – задокументована інформація, що виникає завдяки науково-технічному розвиткові суспільства. Вона потрібна керівникам, працівникам наукової, інженерної і технічної галузей у процесі їхньої діяльності. Сюди належить також спеціалізована економічна та нормативно-правова інформація.

Інформацію можна вважати науковою, якщо вона відповідає низці вимог. Звернемося до визначення: оскільки її одержують у процесі наукових досліджень, то вона нерозривно пов'язана з виробничою, практичною діяльністю людини. З іншого боку – це логічна інформація, що виникає методом обробки інформації, яку одержує людина за допомогою органів чуттів, абстрактно-логічного мислення. Наприклад, сукупність даних про температуру в різних частинах країни ще не є науковою інформацією. Вона набере наукового характеру, якщо буде встановлено

взаємозв'язки між цими частинами. Потрібно взяти до уваги й третю умову – інформація має адекватно відображати об'єктивний світ. Четверта неодмінна умова – вона має використовуватися в суспільно-історичній практиці. Це означає, що, наприклад, наукова фантастика не може належати до наукової інформації, також не можна вважати науковою адекватну та логічну інформацію про коливання температури на березі моря, коли вирішується питання про поїздку у відпустку. Але, якщо ці дані буде оброблено задля створення прогнозу погоди, то вони вже будуть науковою інформацією.

2.1. Властивості наукової інформації

Наукова інформація, щоб її вважали такою, повинна відповідати низці вимог. З погляду хімії, найважливіші властивості наукової інформації – це її достовірність та відтворюваність.

Достовірність інформації (англ. Validity of information) – властивість інформації бути правильно сприйнятою.

Загалом достовірність інформації досягається:
– зазначенням часу подій, про які йдеться;
– зіставленням даних, одержаних з різних джерел;
– своєчасним відкриттям дезінформації;
– усуненням хибної інформації та ін.

Відтворюваність результатів вимірювань (англ. Reproducibility of measurement) – повторюваність (у межах встановленої похибки) результатів вимірювань однієї і тієї ж величини, отриманих у різних місцях, окремими методами, засобами, операторами, в різний час, але зведених до одних і тих самих умов вимірювань.

Відтворюваності інформації досягають завдяки використанню точних приписів, методик проведення експериментів, багаторазовим повторенням дослідів.

Ще однією важливою властивістю наукової інформації є її кумулятивність.

Кумулятивність наукової інформації – здатність наукової інформації набувати більш точного, узагальненого і компактного

вигляду в процесі створення нової наукової інформації.

Кумулятивність є наслідком науково-технічного прогресу. З кумулятивністю наукової інформації тісно пов'язано таке явище, як її старіння.

Старіння наукової інформації – втрата інформацією практичної цінності для споживача внаслідок її кумулятивності або зміни самого описаного об'єкта.

Ступінь старіння документальної інформації залежить від її виду і не є однаковою для різних видів документів. Розглянемо це явище більш детально.

2.2. Старіння інформації

Старіння в нашому світі – це природний процес. Стосовно документів старіння тлумачиться не як фізичне старіння носія інформації, а як досить складний процес старіння інформації, що міститься в ньому. Зовні цей процес виявляється у втраті вченими та фахівцями інтересу до публікацій зі збільшенням часу, що минув з дня їх видання.

Приклад 1. *Запити іноземних наукових журналів у центральній бібліотеці Пущинського центру РАН за 2006 рік [28]:*

2001–2006 – 70,7 %;
1981–2000 – 19,8 %;
1961–1980 – 7,5 %;
до 1960 – 2 %.

До тих публікацій, що вийшли порівняно давно, звертаються набагато рідше, що дає змогу стверджувати факт їх старіння. Які ж механізми керують процесом старіння документів?

Один із них безпосередньо пов'язаний з кумуляцією, агрегацією наукової інформації. Часто матеріал, на викладення якого сто років тому був потрібен цілий курс лекцій, тепер можна пояснити за декілька хвилин за допомогою двох–трьох формул. Відповідні курси лекцій безнадійно старіють: їх ніхто вже не використовує.

Після отримання точніших значень старіють приблизні дані,

а отже, – і документи, у яких вони опубліковані. Тому, коли говорять про старіння наукової інформації, найчастіше мають на увазі саме її уточнення, більш конкретний, стислий і узагальнений виклад у процесі створення нової наукової інформації. Це можливо завдяки тому, що наукова інформація має властивість кумулятивності, тобто допускає коротший, узагальнений виклад.

Іноді старіння документальної інформації має інший механізм: об'єкт, опис якого ми маємо у своєму розпорядженні, з часом змінюється настільки, що інформація про нього стає неточною. Так старіють географічні карти: на місці пустель з'являються пасовища, виникають нові міста й водосховища.

Процес старіння можна розглядати і як втрату інформацією практичної корисності для споживача. Це означає, що він уже не може її використовувати для досягнення певної мети.

І, нарешті, цей процес може бути розглянутий з позицій зміни тезауруса людини. Тобто одна і та ж інформація може бути "застарілою" для однієї людини і "не застарілою" – для іншої.

Ступінь старіння документальної інформації неоднаковий для різних видів документів. На швидкість її старіння впливають різною мірою дуже багато чинників. Особливості старіння інформації в кожній із галузей науки і техніки не можуть бути виведені на основі абстрактних міркувань або усереднених даних статистики – вони органічно пов'язані з тенденціями окремого розвитку кожної з них.

Для того, щоб якось кількісно оцінити швидкість старіння інформації, бібліотекар Р. Бартон (R. E. Burton) і фізик Р. Кеблер (R. W. Kebler) зі США за аналогією з періодом напіврозпаду радіоактивних речовин увели "напівперіоди життя" наукових статей. Напівперіод життя – це час, протягом якого була опублікована половина всієї використовуваної в певний час літератури однієї з галузей знань або предмета. Якщо напівперіод життя публікацій з фізики дорівнює 4,6 року, то це означає, що 50 % усіх нині використовуваних (цитованих) публікацій з цієї галузі мають вік не більше 4,6 року. Напівперіоди життя кожної з галузей науки різні. Бартон і Кеблер наводять такі дані: для публікацій з фізики – 4,6, фізіології – 7,2, хімії – 8,1, ботаніки – 10,0, математики – 10,5, геології – 11,8 року. Проте хоча властивість старіння інформації і

має об'єктивний характер, та вона не розкриває внутрішнього процесу розвитку цієї ділянки знань і має швидше описовий характер. Тому до висновків про старіння інформації варто ставитися дуже обережно.

Проте навіть приблизна оцінка швидкості старіння інформації та документів, що її містять, має величезну практичну цінність: вона допомагає тримати в полі зору тільки ту частину документів, у якій, найімовірніше, розміщена основна інформація про цю науку. Це важливо не тільки для працівників науково-технічних бібліотек й органів науково-технічної інформації, а й власне для споживачів цієї інформації.

2.3. Питання для самоперевірки

1. Дайте визначення інформації.
2. Що таке наукова інформація?
3. Що таке науково-технічна інформація?
4. Які властивості наукової інформації?
5. Що таке достовірність?
6. Що таке відтворюваність?
7. Що таке кумулятивність?
8. Чому старіє наукова інформація?
9. Що таке "напівперіод життя" інформації?

3. СТРУКТУРА ДЖЕРЕЛ НАУКОВОЇ ІНФОРМАЦІЇ

Наукову та науково-технічну інформацію містять публікації включно зі статтями й тезами, монографіями, авторефератами і дисертаціями, рефератами і анотаціями тощо.

У зв'язку з кумулятивністю наукової інформації нові знання потрібно публікувати якомога швидше. Навіть найбільш досконале й цікаве дослідження, якщо воно не опубліковане, не дає ніякої користі для розвитку науки. Рано чи пізно воно втратить свою актуальність і цікавитиме хіба істориків, котрі розглядатимуть сумну долю винаходів і теоретичних концепцій, які свого часу могли дати поштовх для розвитку науки. Тому священний обов'язок ученого – публікувати результати своїх досліджень.

Наукова публікація – це певним способом оформлене джерело

наукової чи науково-технічної інформації, доступне для масового ознайомлення.

Наукові публікації можуть мати різне призначення щодо інформації, що в них міститься. Всі наукові публікації за Потаповим і Кочетовою [22] можна умовно розділити на чотири типи або рівні.

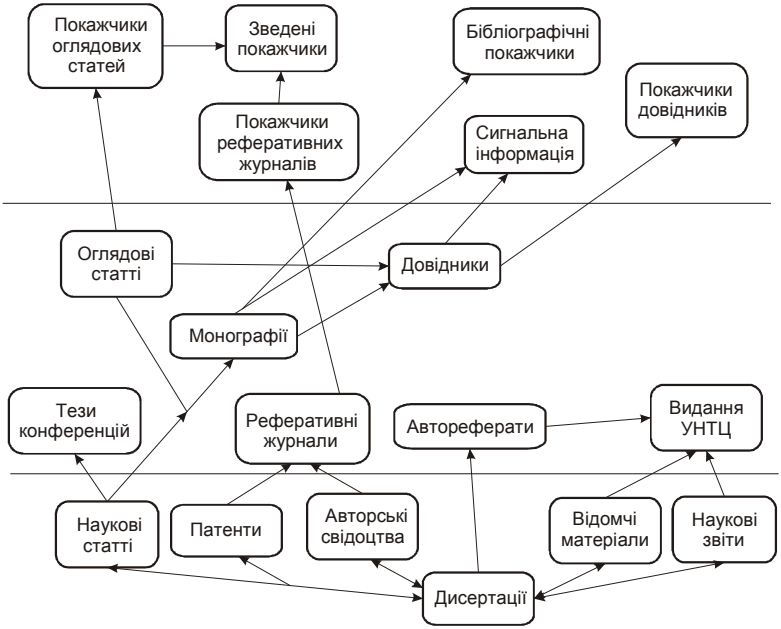


Рис. 1. *Структура джерел наукової інформації*

Первинні джерела містять результати наукових досліджень і фак.-тичний матеріал. Найбільш поширеними є наукові статті й патенти.

У вторинних джерелах зазвичай подають аналіз первинної інформації або певну інформацію про неї та наводять посилання на першоджерела наукової інформації.

Третинні джерела – покажчики та індекси. За їх допомогою можна визначити, в якому джерелі шукати певну конкретну інформацію.

Джерела інформації четвертого рівня подібні до джерел

третього за винятком того, що містять зазвичай інформацію зі всіх випусків покажчиків за попередні роки. Ми в цій схемі обмежимося трьома рівнями.

4. ДРУКОВАНІ НАУКОВІ ПУБЛІКАЦІЇ

Не зважаючи на бурхливий розвиток електронної техніки, більшість джерел наукової інформації подано у друкованому вигляді.

4.1. Первинні джерела наукової інформації

Як уже було сказано, результати безпосередніх досліджень учених з'являються насамперед у первинних джерелах, до яких належать: наукові статті, що публікують в наукових журналах та збірниках, патенти, авторські свідоцтва, наукові звіти, дисертації, матеріали досліджень комерційних компаній та державних установ. Як правило, вони містять найбільш цінну інформацію – опис методик проведення експериментів, що дозволяє відтворити їх у лабораторії або проводити аналіз одержаних даних з того чи іншого погляду.

Найбільш доступними джерелами є наукові статті, менш доступні – патенти, звіти. Дисертації у нашій країні малодоступні, а матеріали відомчих досліджень часто є секретними й недоступними.

4.1.1. Наукові журнали

Розвиток журналів як особливої форми видання триває і нині. Їх поділяють на: вузькоспеціалізовані й багатодисциплінарні; національні та міжнародні; авторитетні й маловідомі. Справжній науковий журнал (journal) від його імітації, а також від науково-популярного журналу (magazine) відрізняє не тільки орієнтація на читацьку аудиторію (включно з професіоналами у відповідній науковій галузі), не лише специфічний "науковий" стиль викладу матеріалу, а й принцип його відбору. Ще в середині XVIII ст. запроваджено систему рецензування (peer review) наукових робіт: статтю, яку подав автор, направляють декільком (як правило – двом) авторитетним у цій галузі фахівцям, які мають засвідчити її новизну, наукову достовірність, а також стиль викладу і відповідність тематики

профілю журналу. Лише після позитивного висновку рецензентів редакція допускає статтю до публікації.

Такий підхід, окрім очевидних переваг (відсівання псевдонаукового і неякісного матеріалу), має й недоліки (не кожне відкриття знаходить розуміння і підтримку в сучасників). Відомо, наприклад, що Ньютон, образившись на колег за недоброзичливе ставлення до його робіт у галузі оптики, взагалі відмовився публікувати свої статті в наукових журналах. Розуміючи недосконаліть системи рецензування, редакції іноді досить поблажливо ставляться до робіт дискусійного характеру і допускають їх до друку, сподіваючись, що істину з часом встановлять інші автори, які зможуть відтворити експеримент в інших умовах або з використанням нових методик.

Читаючи наукову статтю, потрібно пам'ятати: факт її публікації сам по собі ще не є доказом того, що все написане в ній є абсолютною істиною. Лише аналізуючи роботи різних авторів, наукових шкіл, що застосовують різні підходи до вивчення певного явища, можна зробити висновок про ступінь достовірності й усієї статті, й окремих її фрагментів.

Ще більшу обережність потрібно проявляти, оцінюючи результати, опубліковані в нерецензованому інформаційному джерелі.

4.1.2. Структура наукового журналу

Кожен науковий журнал має власну унікальну назву (Publication Title, Publication, Title), в якій зазвичай вказано його тематику і/або найменування організації-видавця.

Приклад 2. *Назви наукових журналів:*

Analytical Chemica Acta;
Journal of Analytical Chemistry;
Журнал аналитической химии;
Український хімічний журнал;
Методи та об'єкти хімічного аналізу.

Оскільки назва журналу може бути довгою, у статтях для зручності використовують стандартні скорочення.

Приклад 3. *Скорочені назви наукових журналів:*

J. Chem. Comm. – Journal of Chemical Communication:
JACS – Journal of American chemical society;
Журн. аналит. хим. – Журнал аналитической химии;
ZAAC – Zeitschrift für anorganische und allgemeine Chemie.

У кожному з журналів правила скорочень можуть суттєво відрізнятися, тому в разі потреби потрібно звертатися до вимог опублікування статей, що наводяться зазвичай на початку чи в кінці номера.

Періодичність виходу журналів варіюється – від тижневиків і двотижневиків до номерів, що виходять раз у декілька років. Зазвичай видання одного року називають томом (Volume) і приписують йому порядковий номер. Кожна книжка журналу в межах тому називається випуском (Issue) і теж має свій порядковий номер.

Як правило, в межах тому застосовують суцільну нумерацію сторінок від першого випуску до останнього; рідше нумерація сторінок починається заново в кожному з випусків. Деякі журнали, що публікують рецензовані та нерецензовані матеріали, використовують складнішу систему.

Приклад 4. *Складна система нумерації сторінок:*

сторінки А1–А10 – новини, реклама;
сторінки 1–146 – наукові статті;
сторінки I–V – анотації книжкових новинок.

Той набір відомостей, за яким конкретну статтю можна відшукати в бібліотеці, називається бібліографічним описом.

4.1.3. Типи публікацій у науковому журналі

У наукових журналах трапляються рецензовані публікації декількох типів, основні з них такі:

– стаття (Article, Paper), в якій досить детально викладено результати закінченої експериментальної або теоретичної наукової роботи (чи її етапу);

– коротке повідомлення (Communication, Short Note) – невелика за розмірами публікація про нові методики, речовини і т. п.; іноді коротке повідомлення називають листом (Letter);

– огляд (Review) – зазвичай досить чимала за обсягом публікація, у якій автор підсумовує власні й чужі досягнення в дослідженні певної тематики.

Деякі журнали спеціалізуються лише на одному з указаних типів публікацій, інші – на декількох.

На підготовку до друку короткого повідомлення потрібно набагато менше часу, ніж для публікацій інших типів, тому, отримавши цікавий результат, автор може опублікувати коротке повідомлення, а пізніше, зібравши більшу кількість матеріалу, запропонувати редакції статтю.

Журнал зазвичай містить й інші тексти, включно з такими, що не пройшли рецензування:

– повідомлення редакції (Editorial) організаційного характеру, що є оглядом анотації матеріалів, опублікованих у цьому випускові;

– листи до редактора (Letter to the Editor), у яких автори висловлюють свою думку про деякі особливо спірні матеріали, опубліковані в журналі (не слід плутати співзвучні, але принципово різні поняття “лист до редактора” і “лист” як синонім “короткого повідомлення”);

– виправлення виявлених помилок, друкарських (Erratum) включно;

– анотації книг (Book Review), комп’ютерних програм (Software Review);

– новини, рекламні матеріали.

Крім того, в журналі розміщують допоміжну інформацію, зокрема:

– зміст (Contents, Table of Contents, TOC) – у кожному випуску;

– авторський покажчик (Author Index), предметний покажчик (Subject Index) або перелік статей у кінці тому;

– правила для авторів (Submission Details, Instructions to Authors) у кожному випуску або в одному з них.

4.2. Інші джерела наукової інформації

4.2.1. Дисертації (Dissertation, Thesis)

Дисертація (від лат. dissertatio – міркування, дослідження) – наукова робота, підготовлена з метою її публічного захисту для

здобуття вченого ступеня.

Дисертаційна робота є докладним звітом про методику та результати певного завершеного дослідження, тому вона є цінним інформаційним джерелом із відповідної теми. Дисертації зазвичай зберігають у бібліотеках тих установ, де проходив захист. Ще по одному примірникові кожної дисертації є в Національній бібліотеці України імені В. І. Вернадського і Українському інституті науково-технічної та економічної інформації. Варто зазначити, що доступ до дисертаційних фондів національних бібліотек досить обмежений.

У нашій країні й інших державах СНД дисертант зобов'язаний видавати автореферат дисертації обсягом у декілька десятків сторінок, в якому коротко викладено головні тези дослідження, що його готують до захисту. Автореферат також розсилають до бібліотек, але зазвичай він належить до матеріалів тимчасового зберігання.

У західних країнах аналог кандидатської дисертації називають Ph.D. Dissertation, Doctoral Thesis і т. п. (категорії, аналогічної вітчизняній докторській дисертації, у більшості країн немає).

Магістерські дисертації (іменовані, наприклад, Master's Thesis) за своїм рівнем можуть бути подібні до наших дипломних робіт, на що потрібно зважати при оцінці матеріалу, розміщеного в них.

4.2.2. Наукові звіти (Research Report)

Кожен науковий колектив, що отримує кошти для проведення науково-дослідної роботи (НДР), зобов'язаний надавати організації, яка його фінансує (фонду, міністерству і т. д.), звіти: проміжні і завершальний. Основні результати НДР зазвичай оприлюднюють у журналах і доповідають про них на конференціях, але у звіті ці ж матеріали подають в узагальненому вигляді. Крім того, при написанні звіту автори не обмежені кількістю сторінок, тому тут можна ознайомитися з такими деталями експерименту, які не згадано в журнальній публікації.

4.2.3. Патенти

Патент – охоронний документ, що засвідчує пріоритет, авторство і право власності на винахід.

Винахід (корисна модель) – результат інтелектуальної діяльності людини в будь-якій сфері технології.

Патент (лат. patere – робити доступним для загалу) засвідчує авторство на винахід та виняткове право на використання його протягом певного строку. Патент видає винахіднику або його правонаступникові Державне патентне відомство. Його дія поширюється тільки на територію держави, у якій цей документ видано. Строк дії патенту встановлюється законодавством (як правило, 15–20 років). Патент може бути оскаржений і анульований на законодавчій основі. З поняттям патенту тісно пов'язаний юридичний термін "патентна чистота", який означає, що машину, прилад, технологічний процес, матеріал, продукт тощо можна використовувати (виготовити, ввезти для продажу) у певній державі без порушення прав патентовласника.

У широкому розумінні "патент" є комплексом виняткових прав на використання винаходу, корисної моделі або промислового зразка, які держава гарантує патентовласнику, тобто винахідникові чи особі, якій винахідник передав виключні майнові права. Виняткові права держава надає на заздалегідь визначений період в обмін на контрольоване та публічне розкриття патентовласником суттєвих ознак винаходу.

Патент надає правовласнику виняткове право заборонити іншим порушувати патент. Лише патентовласник може надати згоду на використання винаходу або передати виняткові майнові права на використання патенту. Однак патент не обов'язково надає винахіднику право застосовувати запатентований винахід. Якщо патент видано на вдосконалення до винаходу, який охороняється патентом, власник цього документа не зможе розпочати виробництво вдосконаленого винаходу, не отримавши дозволу від власника патенту на оригінальний винахід.

У деяких країнах існує низка законодавчих актів, які дозволяють державі у випадках, коли винахід має особливо важливе значення для країни, але з патентовласником не

досягнуто угоди про видачу ліцензії або передачу патенту, примусово викуповувати патент з виплатою винагороди власникові. Як спосіб захисту патентовласника, законодавством багатьох країн передбачено штрафні санкцій при ввезенні на територію країни виробів, у яких втілено аналогічний винахід (штрафи, накладання арешту на ввезені товари, пред'явлення позову про відшкодування збитків тощо).

4.2.3.1. Види патентів

Є декілька видів патентів, що відрізняються один від одного тільки сутністю та ретельністю проведеної патентної експертизи і, відповідно, заголовком. Згідно із законом України № 3687-12 "Про охорону прав на винаходи і корисні моделі" (за останньою редакцією від 07.07.2009 р.) – це патент на винахід, деклараційний патент на винахід, деклараційний патент на корисну модель, деклараційний патент на секретний винахід, деклараційний патент на секретну корисну модель.

Патент на винахід – різновид патенту, що видається за результатами кваліфікаційної експертизи заявки на винахід.

Деклараційний патент на винахід – різновид патенту, що видається за результатами формальної експертизи заявки на винахід.

Деклараційний патент на корисну модель – різновид патенту, що видається за результатами формальної експертизи заявки на корисну модель.

Патент (деклараційний патент) на секретний винахід – різновид патенту, що видається на винахід, віднесений до державної таємниці.

Деклараційний патент на секретну корисну модель – різновид патенту, що видається на корисну модель, віднесену до державної таємниці.

4.2.3.2. Патентоздатність

Патентоздатність – сукупність ознак технічного рішення, які відзначаються новизною і є неодмінними та достатніми для визнання його винаходом.

Оскільки патент надає його власнику обмежену монополію на використання об'єкта промислової власності, щодо якого він виданий, суспільство зацікавлене в тому, щоб таку монополію отримували лише ті об'єкти, які є дійсно корисними, і лише ті заявники, які насправді створили ці об'єкти. Для цього призначено державну експертизу, у процесі якої з'ясовується відповідність заявленого об'єкта до умов надання правової охорони. У патентному праві ці умови отримали назву "критеріїв патентоздатності". Якщо в результаті експертизи встановлюють відповідність заявленого об'єкта до критеріїв патентоздатності, такий об'єкт вносять до Державного реєстру і видають на нього правопідтверджувальний документ, яким і є патент.

4.2.3.3. Патентна формула

Патентна формула (англ. Claims) – коротке формулювання об'єкта патентного захисту, яким закінчується патентний опис. На основі патентної формули визначають обсяг прав патентовласника.

4.2.3.4. Територіальний характер патенту

Основою патентного права є принцип територіальності, згідно з яким правова охорона винаходу надається лише у межах території країни, компетентне відомство котрої здійснило експертизу патентоздатності. Територіальний характер дії патенту скасовується за умови укладання міжнародних угод, наприклад Конвенції про єдине патентне право країн ЄС.

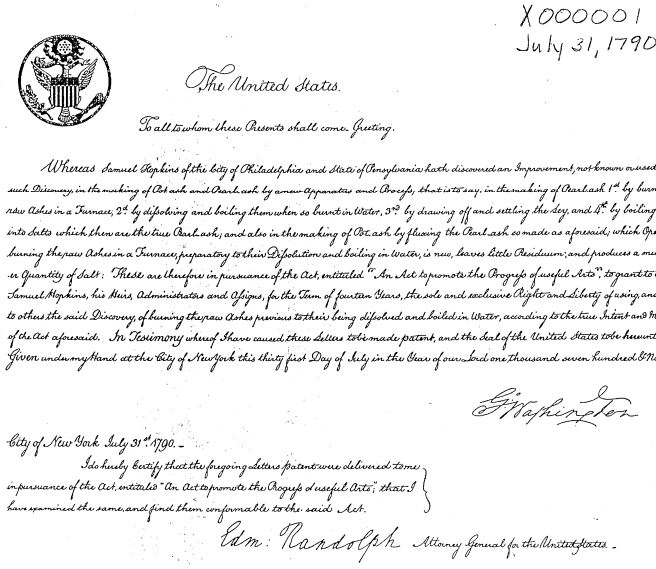
4.2.3.5. Реєстрація патенту

Патент отримують через подання до відповідного патентного відомства заявки на патент, у якій має міститися опис винаходу та, в окремих юрисдикціях, докази його корисності.

В Україні Реєстр патентної інформації є офіційним

документом, призначеним для реєстрації громадян країни, які атестовані як представники у справах інтелектуальної власності (патентні повірені), а також внесення усіх змін щодо їхньої реєстрації. Реєстр патентної інформації – це зброшуровані, прошнуровані та скріплені печаткою книги, які, разом з його копією на магнітних носіях, підлягають постійному зберіганню в Державному департаменті інтелектуальної власності, що діє у складі Міністерства освіти і науки України. Цей орган забезпечує також здійснення реєстрації та інших дій, пов'язаних з веденням Реєстру.

Реєстрація патентних повірених здійснюється згідно з Положенням про представників у справах інтелектуальної власності (патентних повірених), затвердженим постановою Кабінету Міністрів України від 10 серпня 1994 р. № 545 (545-94-п), на підставі рішення атестаційної комісії про атестацію патентного повіреного та наказу голови Департаменту про внесення патентного повіреного до Державного реєстру.



X000001
July 31, 1790

The United States.

To all to whom these Presents shall come. Greeting.

Whereas Samuel Hopkins of the City of Philadelphia and State of Pensylvania hath discovered an Improvement, not known or used before such Discovery, in the making of Pot ash and Pearl ash by new Apparatus and Process, that is to say, in the making of Pearl ash 1st by burning the raw Ashes in a Furnace, 2d by dissolving and boiling them when so burnt in Water, 3rd by drawing off and settling the ley, and 4th by boiling the ley into salts which then are the true Pearl ash, and also in the making of Pot ash, by fluxing the Pearl ash so made as aforesaid; which Operation of burning the raw Ashes in a Furnace preparatory to their Dissolution and boiling in Water, is new, leaves little Residuum, and produces a much greater Quantity of Salt: These are therefore in pursuance of the Act, entituled "An Act to promote the Progress of useful Arts", to grant to the said Samuel Hopkins, his Heirs, Administrators and Assigns, for the Term of fourteen Years, the sole and exclusive Right and Liberty of using, and vending to others the said Discovery, of burning the raw Ashes previous to their being dissolved and boiled in Water, according to the true Intent and Meaning of the Act aforesaid. In Testimony whereof I have caused these Letters to be made patent, and the Seal of the United States to be hereunto affixed. Given under my Hand at the City of New York this thirty first Day of July in the Year of our Lord one thousand seven hundred & Ninety.

G. Washington

City of New York July 31st 1790.

I do hereby Certify that the foregoing Letters patent were delivered to me in pursuance of the Act, entituled "An Act to promote the Progress of useful Arts", that I have examined the same, and find them conformable to the said Act.

Edm: Randolph Attorney General for the United States.

Рис. 2. *Перший патент США*

У процесі ведення Реєстру до нього вносять зміни та доповнення до інформації про патентних повірених. Підставою для внесення змін та доповнень до Реєстру є рішення голови Департаменту.

Відомості, що належать до реєстрації патентних повірених, а також відомості про зміни та доповнення, що занесені до Реєстру, Департамент публікує у своєму офіційному бюлетені.

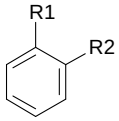
Після внесення до Реєстру відомостей щодо реєстрації патентних повірених Департамент може надати витяг з Реєстру за заявою будь-якої особи. Витяг стосується даних про реєстрацію одного патентного повіреного.

Перший в історії незалежної України патент на винахід видано Державним патентним відомством України 30 листопада 1992 р. Він охороняв спосіб отримання антифрикційного матеріалу для поверхонь тертя. Власником патенту став Микола Костянтинович Ященко. На 2010 р. цей документ вже не є чинним.

Уперше патенти з'явилися в 1474 р. у Венеціанській Республіці, у 1623 р. – в Англії, 1790 р. – у США, 1812 р. – на території царської Росії.

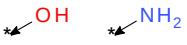
4.2.3.6. Патентування хімічних сполук

Для патентування хімічних сполук розроблено так звані структури Маркуша.



R1 = 

R1 >0

R2 = 

R2 >0

Рис. 3. *Приклад структури Маркуша*

Формула Маркуша сама по собі не відповідає ніякій конкретній сполуці, вона є лише зручним способом позначення хімічних структур в узагальненому вигляді. На практиці їй може відповідати велика кількість конкретних сполук. До формули Маркуша входить один або декілька варіантів замісників, що об'єднані в групу альтернативних структур. Вона складається із загальної, кореневої структури та R-груп, що описують можливі замісники.

Цій структурі Маркуша будуть відповідати такі структурні формули:

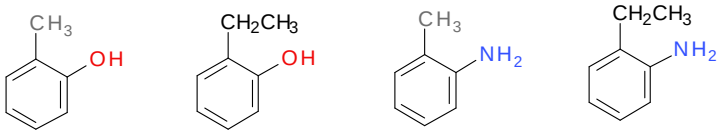


Рис. 4. *Приклад структур, що їх можна одержати з родоначальної структури Маркуша*

4.2.4. Авторські свідоцтва

Авторське свідоцтво – документ, що засвідчує юридичне право автора. Найбільш поширеним був у СРСР. Автори, які працювали на підприємствах, у науково-дослідних установах та інших організаціях, подавали заявки на одержання авторського свідоцтва, як правило, через своє підприємство чи організацію. Якщо роботу було виконано як службове завдання, керівники підприємств і організацій оформляли заявки на видачу авторського свідоцтва на ім'я підприємства (організації) із зазначенням автора (авторів). Якщо винахід був результатом колективної творчості і при цьому неможливо встановити авторство окремих осіб, авторське свідоцтво видавали на ім'я того підприємства (організації), де працює автор. За автором зберігали повні особисті та майнові права, а також пільги (право авторства, право на надання імені автора, права на першочергове здобуття посад наукових працівників за своїм

фахом тощо). В Україні авторське свідоцтво видавав Комітет у справах винаходів та відкриттів при Кабінеті Міністрів України.

Сьогодні авторське свідоцтво замінено патентом на винахід та корисну модель.

4.2.5 Відомчі матеріали

Багато з приватних та державних компаній мають власні науково-дослідницькі підрозділи, які проводять і фундаментальні, і прикладні дослідження. Але їхні результати зазвичай являють собою комерційну таємницю й не підлягають розголошенню. Хоча трапляється, що матеріали досліджень публікують у засобах масової інформації. Найчастіше результати досліджень представлено у вигляді патентів та вторинних джерел наукової інформації. Звичайно, для підтримки престижу співробітники компаній часто публікують такі матеріали також у відкритій літературі.

4.3. Вторинні джерела інформації

Вторинні джерела наукової інформації містять дані про первинну наукову інформацію. Сюди належать: анонси (автореферати дисертацій, препринти, тези конференцій); аналітичні видання (оглядові статті, монографії, реферативні видання та збірники фактичної інформації – довідники).

4.3.1. Сигнальна інформація

Сигнальна інформація – вид інформаційного обслуговування, що виконує функцію попереднього сповіщення про документи, публікуючи їх бібліографічні дані. Прототипом сигнальної інформації слугувала бібліографічні списки нових надходжень у бібліотеках, у яких подавали перелік книг, періодичних видань, що продовжуються у монографіях. На відміну від них, сигнальна інформація, яку почали застосовувати з 50-х рр. XX ст., дає відомості переважно про статті в науково-технічних журналах і збірках. Цю інформацію можна поширювати і вибірково, й у вигляді інформаційних видань (систематизованих бюлетенях або покажчиках змісту журналів). Найбільш відомими бюлетенями сигнальної інформації, що випускаються за допомогою автоматизованих систем, є: сигнальна інформація російського ВІНІТІ з автоматики,

радіоелектроніки, хімії, металургії, фізики, біології та інформатики (понад 70 випусків охоплюють близько 400 тис. публікацій за рік), “Current Contents” Інституту наукової інформації США (5 серій, близько 300 тис. публікацій) і “Chemical Titles” Американського хімічного товариства.

4.3.2. Анонси

Анонси можна було б віднести до рефератів, але вони описують не опубліковану інформацію, а ту, що тільки має вийти друком або буде офіційно представлена. Сюди належать препринти, тези конференцій, автореферати дисертацій тощо.

4.3.3. Препринти

Процес видання статті в науковому журналі складається з багатьох стадій (первинний аналіз рукопису, розсилання її рецензентам, рецензування, листування з автором і коректування тексту, підготовка рукопису для друку, друкування, розсилання передплатникам) і в докомп’ютерну епоху він тривав протягом декількох місяців. Для того, щоб прискорити розповсюдження інформації, в другій половині XX ст. було винайдено нову форму публікації – препринт (preprint).

За своєю структурою друкарський препринт повністю відповідає науковій статті, але його розповсюджують у вигляді окремої брошури між обмеженою кількістю колег автора. Головна відмінність препринта полягає у тому, що він не підлягає рецензуванню; це лише “напівфабрикат”, який подаватимуть до публікації в традиційному періодичному виданні.

У представників різних наук неоднозначне ставлення до препринтів: від схвалення серед фізиків-ядерників до абсолютного заперечення медиками, які вважають неприпустимим поширення нерецензованої інформації. Хіміки почали виявляти цікавість до препринтів лише протягом останніх десятиліть.

Примітка. Співзвучні з назвою препринт, але принципово відрізняються публікації, звані репринт, оффпринт, постпринт. Це повторні видання опублікованих раніше статей або їх фрагментів, тобто матеріалу, що пройшов рецензування.

Окрім періодичних видань – журналів, існують продовжувані видання, що виходять у міру накопичення матеріалу, без

строгої періодичності, але з нумерованими й однотипно оформленими випусками. Звичайно вони називаються “Праці”, “Вчені записки”, “Бюлетень”, “Збірка” і видають їх наукові організації та університети. Статті, що публікуються в таких збірниках, проходять стадію рецензування, але зазвичай лише в межах тієї організації, яка є власником відповідного видання.

Для публікації наукових робіт, що мали значення тільки для вузького кола спеціалістів, наприкінці XX ст. широко використовувалася система депонування рукописів. Рукопис, що депонується, як і рукопис звичайної статті, проходив весь цикл рецензування і, при потребі, коректування змісту, але його не публікували в журналі, а заносили на зберігання (депонування) в один зі спеціальних центрів. Організація, що займалася депонуванням (Український інститут наукової і технічної інформації, редакція журналу або наукова організація), публікувала список отриманих матеріалів, а також реферати статей. Особи, зацікавлені в ознайомленні з повним текстом рукопису, могли замовити й отримати копію такої роботи.

Значний пласт наукової і науково-технічної інформації міститься також у патентах.

І нарешті, про термінологію. В англомовній літературі трапляється словосполучення Grey Literature, якому немає відповідника в українській мові. Так позначають ту частину наукової, науково-технічної і ділової інформації, яка поширюється не по каналах традиційних комерційних видавництв. Grey Literature містить, зокрема, всі форми наукових звітів, дисертації, малотиражні матеріали конференцій, технічну документацію, препринти і т. д.

4.3.4. Матеріали конференцій

У науковому світі конференція виконує декілька функцій. По-перше, це засіб безпосереднього спілкування вчених, можливість обмінюватися ідеями, досягненнями, попередніми результатами досліджень. Зазвичай, конференції є вузькоспеціалізованими, й на них збираються фахівці певного напряму науки. Це дає змогу вченому бути в курсі основних досягнень у

цій галузі, а у процесі обговорення їх результати проходять попереднє неформальне рецензування.

Видатний радянський вчений П. Л. Капіца високо оцінив роль безпосереднього спілкування науковців. У своїй доповіді на засіданні АН СРСР, присвяченому 250-річчю з дня народження М. В. Ломоносова, він зазначав: “Схема, за якою відбувається інтернаціональна співдружність учених, добре відома і залишається зараз такою ж, якою вона була за часів Ломоносова. У різних країнах є свої групи науковців, які працюють при університетах, академіях або інших наукових інститутах. Оскільки кожна наукова область або проблема може розвиватися лише в одному напрямі, то щоб не збиватися з цього істинного шляху, доводиться повільно рухатися і витрачати багато сил на пошукові роботи. Співпраця в науковій роботі розподіляється між колективами вчених, що працюють над певним питанням. Роботи вченого, що проходять поза колективом, звичайно залишаються непоміченими. Життя незмінно показує, що така колективна робота вчених як усередині країни, так і в міжнародному масштабі можлива лише при особистому контакті. Ученому для того, щоб його наукову роботу визнали, потрібно не лише її опублікувати, а він ще має переконати людей в її правильності й довести її значення. Все це успішно можна зробити лише при особистому контакті. Як за часів Ломоносова, так і в наш час, щоб учений своїми роботами міг впливати на колективну роботу, потрібне особисте спілкування, потрібен живий обмін думками, потрібна дискусія, всього цього не може замінити ні друкарська робота, ні листування. ...” [18, 181]. Саме цим П. Л. Капіца пояснює той факт, що роботи таких видатних російських вчених, як Ломоносов та Петров свого часу не були належно визнані. Тому наукові конференції є однією з важливих форм роботи науковця, і в них потрібно брати участь.

Учений, що хоче взяти участь у конференції, надає оргкомітету тему і тези майбутнього повідомлення. Як правило, перед конференцією організатори видають невеликим тиражем збірку тез доповідей (Abstracts), що входять до програми засідань.

Інформаційну цінність тез доповідей складно охарактеризувати однозначно. З одного боку, за змістом матеріалів збірки можна чітко визначити, над якими проблемами працюють колеги і приблизно зрозуміти, які підходи при цьому використовують. З іншого боку, тези мають невеликий обсяг (зазвичай це одна сторінка тексту), складаються головно із загальних положень і не завжди за змістом збігаються з тією доповіддю, яка прозвучить на засіданні.

Більше інформації містять розширені тези (Enhanced Abstracts), які розробляють на деяких конференціях – вони і за обсягом більші, і структурно подібні до коротких статей.

Досить часто після закінчення конференції праці (Proceedings), відкоректовані з урахуванням дискусій, що були спричинені виступами, публікують окремою книгою або на сторінках періодичного видання.

4.3.5. Реферативні журнали

У світі публікують так багато наукових статей, що жоден учений не спроможний перечитувати всі першоджерела навіть зі своєї спеціальності. Потребу створення допоміжних інформаційних засобів усвідомили ще в XIX ст., і 1830 р. в Лейпцигу з'явився перший хімічний реферативний журнал "Pharmaceutisches Central-Blatt" (з 1856 р. – "Chemische Zentralblatt").

Реферативний журнал містить бібліографічні описи й реферати наукових публікацій (статей, книг, тез доповідей тощо). Матеріал у ньому розподілено за тематичними розділами і підрозділами, що дає змогу читачеві досить швидко скласти думку про літературні новинки в галузі, яка його цікавить, і ефективно сформувати для себе список тих робіт, які потрібно детальніше вивчити з першоджерела.

Сьогодні наукове співтовариство видає два універсальні загальнохімічні реферативні журнали: реферативний журнал "Хімія" (Росія) і "Chemical Abstracts" (США), а також значну кількість спеціалізованих, наприклад: "Analytical Abstracts" (Великобританія), "Джерело" (Україна). Друковані варіанти реферативних журналів є об'ємними фоліантами, що містять

реферати, (як-от РЖ "Химия" – 24 випуски за рік), а також не менші за обсягом книги з покажчиками (Index): авторським, предметним, формульним й ін. У докомп'ютерну еру покажчики були єдиним інструментом, що дав змогу проводити інформаційний пошук у журналі.

На рис. 5 і рис. 6 наведено приклади рефератів із реферативних журналів.

15 Г9. Хелатные комплексы металлов с 2-окси-5-метилацетофенон-(ос)нафтилом. Joshi G. B., Patel G. S., Thakor V. M. Metal chelate complexes of 2-hydroxy-5-methyl-acetophenone-(oc)naphthil. «Chem. Era», 1976, **12**, № 6, 202—203 (англ.)

Описано получение и изучены аналитич. св-ва хелатов Cu, Ni и Co с 2-окси-5-метилацетофенон-(ос)нафтилом (**I**). Показано, что Cu, Ni и Co при pH 10,0—10,8, 9,0—9,8 и 10,8—11,5 *образуют* с **I** *красновато-кор., красн.* и зел. осадки соотв. с отношением компонентов металл : : **I** = 1 : 2. Комплексы Cu и Co с **I** парамагнитны, Ni — диамагнитны. Изучены абсорбц. спектры р-ров хелатов металлов в $CHCl_3$ в видимой и УФ-областях. Показана возможность гравиметрич. определения Cu в присутствии Ni и Co при pH 10,8. О. П. Елисеева

Рис. 5. *Зразок реферату з реферативного журналу "Химия"*

01B29. **Characterization of chromatographic carrier materials.** Kueppers, H.-J.; Zilse, M.; Thommes, M.; Kothmayr, T. (Physikalische Analytik, Sachtleben Chemie GmbH, 47198 Duisburg, Germany). *LaborPraxis*, Jun 2001, **25** (6), 74-76, 78 (German).—Specific surface areas, pore volumes, pore sizes and pore size distributions were measured by gas sorption with liquid N_2 or mercury porosimetry, density by gas or liquid pycnometry, particle size and particle size distribution by laser granulometry and morphology by SEM with digital image analysis. These procedures were applied to five porous TiO_2 products (Sachtopore, Sachtleben) recently introduced for LC.

Рис. 6. *Зразок реферату з реферативного журналу "Analytical Abstract"*

Як бачимо, реферат із реферативного журналу "Химия" має більш інформативний характер і в деяких випадках може

замінити оригінальну статтю. На відміну від нього реферат з “Analytical Abstract” має більш загальний характер і в ньому не описано конкретних практичних результатів.

4.4. Третинні джерела наукової інформації

Третинними джерелами інформації є покажчики (або індекси) – бібліографічні, покажчики оглядових статей, покажчики довідників та найбільш поширені – покажчики реферативних журналів: предметні, авторські, формульні тощо. По суті, вони схожі на словники, в яких терміни розміщено в алфавітному порядку і вказано посилання на конкретні номери рефератів або інших інформаційних джерел. Потапов і Кочетова [22] виділяють ще четвертий рівень інформаційних джерел – зведені покажчики (англ. Cumulative Indexes), які відрізняються від джерел третього рівня тільки тим, що охоплюють певну кількість випусків покажчиків, виданих протягом певного періоду, зазвичай за декілька років.

За своїми характеристиками третинні джерела інформації найбільш пригодні до внесення в електронні бази даних, тому їх найбільш інтенсивно переносять на електронні носії інформації.

4.5. Питання для самоперевірки

1. Що таке первинні джерела наукової інформації?
2. Дайте визначення вторинних джерел наукової інформації.
3. Що називають третинними джерела наукової інформації?
4. У чому полягають основні відмінності між різними рівнями джерел наукової інформації?
5. Для чого видають реферативні журнали?
6. У чому полягає важливість наукових конференцій?
7. Що таке патент?
8. Що таке препринт?

5. СТРУКТУРА ДЖЕРЕЛ ПЕРВИННОЇ НАУКОВОЇ ІНФОРМАЦІЇ

5.1. Структура наукової статті

Структуру сучасних наукових публікацій остаточно було

сформовано до початку XX ст. Більшість первинних джерел наукової інформації мають багато спільного за своєю будовою; їх основними елементами є назва, відомості про авторів, реферат, основний текст, список використаних джерел.

5.1.1. Назва (Title)

У природничо-наукових журналах, як правило, формулюють назву за можливості стисло і так, щоб вона відображала зміст конкретної публікації. Це дуже корисне правило – воно дозволяє читачеві відбирати матеріал, що цікавить його, користуючись лише змістом.

У журналах гуманітарного профілю досить часто можуть траплятися назви помітні, але малоінформативні. Хто б, наприклад, міг припустити, що у статті «Великі надії: новий підхід у навчанні» йдеться всього лише про введення до навчальних програм нового курсу з бухгалтерської справи?

5.1.2. Відомості про авторів

Нижче (іноді вище) назви статті вказують усіх авторів (Author) публікації. У різних країнах та в різних журналах існують власні традиції відображення персональних даних: ім’я і прізвище, ініціали і прізвище та ін. Варіанти написання персональних даних одного й того ж автора відрізняються (наприклад: J. F. Smith, John F. Smith, John Fitzgerald Smith), і це обов’язково потрібно взяти до уваги, проводячи вузькоспрямований інформаційний пошук.

У більшості випадків указують місце роботи авторів (Affiliation), а також адреси для листування (Correspondence Address).

5.1.3. Реферат або анотація (Abstract, Summary, Resume)

У рефераті у стислій формі окреслюють суть публікації: що саме досліджували і які основні результати отримано.

Проводячи інформаційний пошук, варто мати на увазі, що для опису одного й того ж матеріалу в рефераті та в основному тексті можуть використовувати різну термінологію. Оскільки реферат має узагальнювальний характер, то в нім велика частка загальних термінів; в основному ж тексті можуть переважати

терміни, що мають вужче значення (наприклад, хлорид, бромід і йодид – в основному тексті, а галогеніди – в рефераті).

Крім того, реферати статей великих комерційних видавництв, що подають наукові статті за окрему плату, часто мають відверто виражений рекламний характер і не містять основних фактичних результатів досліджень. Якщо раніше студент міг написати огляд до курсової роботи за результатами аналізу лише рефератів статей, то останнім часом зробити це майже неможливо.

5.1.4. Основний текст

Структура тексту статті залежить від того, яка це робота – теоретична, експериментальна, прикладна, методологічна, методична і т. д. Проте у будь-якому разі у явній або неявній формі у цій частині статті має бути огляд літератури й обґрунтування мети, опис методики дослідження, опис отриманих результатів, ілюстрації, обговорення результатів, висновки.

Подібну будову мають і короткі повідомлення. В оглядах же переважає структуризація за тематичною ознакою.

Основний текст експериментальних статей зазвичай складається з короткого вступу – огляду літератури, опису використаних реактивів, приладів та режимів їхньої роботи, обговорення одержаних результатів, опису методик.

5.1.5. Посилання і список використаної літератури (References)

Здавна у науковій літературі взяли за правило вказувати, що зробили інші дослідники з обговорюваної тематики. По-перше, так автор чітко окреслює свій особистий внесок, а наукове співтовариство як саморегульована система оцінює якість та обсяг роботи кожного з учасників. По-друге, це допомагає читачеві орієнтуватися в інформаційному просторі – після ознайомлення зі статтею знаходити інші наукові роботи, тематично пов’язані з нею.

У тих частинах статті, де автор цитує чужі дані або вважає за потрібне згадати роботу попередника, поміщають посилання – порядковий номер, який розшифровують у списку використаних джерел, що розміщується звичайно в кінці статті.

Приклад 5. *Фрагмент тексту з посиланнями:*

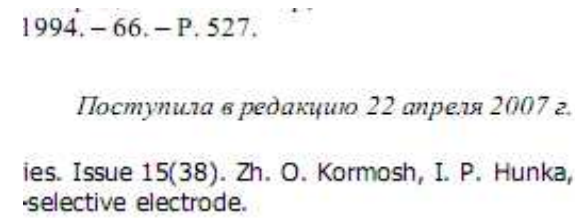
Для визначення піроксикаму використовують потенціометричні [2], спектрофотометричні [3–6], флуориметричні [7] та інші методи.

У гуманітарних журналах використовують й іншу форму посилання – не порядкові номери, а прізвища авторів, на яких посилаються в конкретному випадку.

У списку використаних джерел (References) указують повні або скорочені бібліографічні описи робіт, на які посилається автор.

5.1.6. Допоміжні елементи статті

Перед статтею або після неї зазначають дату отримання рукопису редакцією (наприклад, Received June 23, 2005) – ці відомості можуть знадобитися при встановленні пріоритету в науковому відкритті.

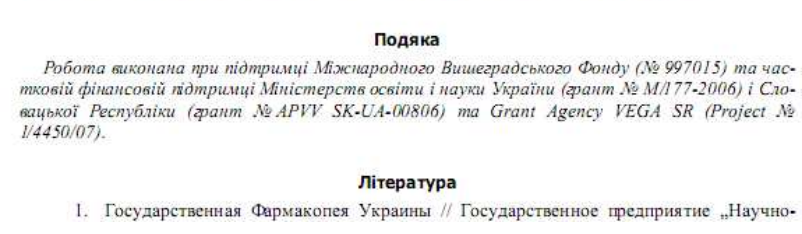

1994. – 66. – P. 527.

Поступила в редакцию 22 апреля 2007 г.

ies. Issue 15(38). Zh. O. Kormosh, I. P. Hunka,
-selective electrode.

Рис. 7. *Дата одержання рукопису редакцією журналу*

Наприкінці статті зазвичай указують організації або осіб, що сприяли виконанню роботи (Acknowledgement).


диклофенаку, яка апробована при його визначенні у різних фармацевтичних препаратах.

Подяка

Робота виконана при підтримці Міжнародного Вишеградського Фонду (№ 997015) та часткової фінансової підтримці Міністерств освіти і науки України (грант № М/177-2006) і Словацької Республіки (грант № APVV SK-UA-00806) та Grant Agency VEGA SR (Project № 1/4450/07).

Література

1. Государственная Фармакопея Украины // Государственное предприятие „Научно-

Рис. 8. *Подяка фондові за сприяння у дослідженнях*

Як окремий елемент стаття може містити додатки (Supplement), наприклад таблиці, схеми.

5.2. Структура патенту на винахід

Патенти мають таку структуру:
– бібліографічні дані;
– назва;
– опис винаходу;
– формула винаходу;
– креслення;
– реферат.


УКРАЇНА UKRAINE

ПАТЕНТ

НА КОРИСНУ МОДЕЛЬ

№ 34773

СПОСІБ ЕКСТРАКЦІЙНО-ФОТОМЕТРИЧНОГО ВИЗНАЧЕННЯ ІНДОМЕТАЦИНУ

Видано відповідно до Закону України "Про охорону прав на винаходи і корисні моделі".

Зареєстровано в Державному реєстрі патентів України на винаходи **26.08.2008.**

Голова Державного департаменту інтелектуальної власності М.В. Паладій

Рис. 9. *Патент України на корисну модель*

5.2.1. Бібліографічні дані

Бібліографічні дані містять відомості, потрібні для реєстрації, зберігання і пошуку патенту: номер патенту; назва країни, що видала патент; дата подачі заявки; дата видачі патенту; класифікаційні індекси (умовні цифрові й буквені позначення розділів систем класифікації винаходів (наприклад, МПК), до

яких належить патент); кількість пунктів патентної формули, ім'я та адреса власника.

5.2.2. Назва

Назва винаходу має бути короткою і точною. Вона, як правило, характеризує його призначення і вказується в однині. Назва є самостійною частиною патенту, оскільки заголовки нерідко перекладають окремо від патентів, складаючи за ними картотеки для пошуку описів винаходів.

5.2.3. Опис винаходу

Опис має розкривати суть винаходу з точністю, достатньою для його здійснення, тобто фахівець зазначеної галузі на підставі опису повинен отримати достатньо інформації для реалізації винаходу.

Опис винаходу містить такі розділи:

– галузь техніки, до якої належить винахід;

– рівень техніки;

– розкриття винаходу;

– короткий, але такий, що відображає головну суть винаходу, опис креслень, схем, малюнків, ескізів (якщо вони містяться в заявці);

– реалізація винаходу.

Якщо до патенту додано креслення, то в повному описі мають розшифровуватися цифри, що позначають на кресленнях деталі пристрою, який патентується.

5.2.4. Формула винаходу

З юридичного погляду формула винаходу призначена для визначення обсягу правової охорони, що її надає патент; у ній сформульовано всі істотні ознаки винаходу.

Формула винаходу складається з одного або декількох пунктів. Кожен пункт зазвичай містить дві частини: обмежувальну і відмітну, розділені словосполученням "що відрізняється тим, що ..." Обмежувальна частина пункту формули містить назву винаходу і його важливі ознаки, вже відомі з рівня техніки. До відмітної частини входять ознаки, складові частини суті винаходу, новизна.

Кожен пункт формули – це одне речення. Пункти формули поділяють на залежні та незалежні. Незалежний пункт формули

винаходу характеризує сукупність його ознак, що визначає обсяг правової охорони, і викладається у вигляді логічного визначення об'єкта винаходу. Залежний пункт формули містить уточнення або розвиток винаходу, розкритого в незалежному пункті.

У патентах англійською мовою патентна формула починається словами: claim, claims (I claim; We claim; What I claim is; What we claim is). Як правило, вона складається з декількох пунктів у вигляді нумерованих абзаців. Якщо у формулі лише один пункт, то його не нумерують. Кожен пункт, яким би містким він не був, в англійських патентах складається з одного речення.

Креслення – це необов'язкова частина патенту. Фігури креслень нумерують і перераховують в описі. Деталі на них позначають цифрами, буквами або іншими індексами, поясненими в описі. Окрім креслень, допускаються схеми, малюнки або інші графічні матеріали, що пояснюють ідею.

Реферат є коротким описом винаходу, який містить опис ознак винаходу (формула) та сфери його застосування.

5.3. Питання для самоперевірки

1. Які головні елементи наукової статті?
2. Компоненти патенту на винахід та корисну модель.
3. Для чого потрібна формула винаходу?
4. Основні пункти опису патенту.

6. БІБЛІОГРАФІЧНИЙ ОПИС

Як уже зазначалося, цитування робіт попередників займає важливе місце у тексті наукової статті. По-перше, цим підкреслюють взаємозв'язок роботи з іншими науковими дослідженнями; по-друге, так визнається статус авторів, що досягли результатів, на які спираються дослідники; по-третє, використання посилань значно полегшує пошук матеріалів з конкретної тематики.

Оскільки бібліографічний опис має важливе значення для проведення ефективної наукової роботи, розглянемо його більш детально.

Бібліографічний опис – сукупність відомостей про документ,

наведених за певними правилами, що встановлюють послідовність областей та елементів, і призначених для ідентифікації та загальної характеристики документа.

Нещодавно уведено новий стандарт з бібліографічного опису ДСТУ ГОСТ 7.1:2006 “Бібліографічний запис. Бібліографічний опис. Загальні вимоги та правила складання”, який замінив п’ять стандартів, створених ще за радянської доби: ГОСТ 7. 1-84 “СИБИД. Библиографическое описание документа. Общие требования и правила составления”; ГОСТ 7. 16-79 “СИБИД. Бібліографическое описание нотних зданий”; ГОСТ 7. 18-79 “СИБИД. Библиографическое описание картографических зданий”; ГОСТ 7. 34-81 “СИБИД. Бібліографическое описание изозданий”; ГОСТ 7. 40-82 “СИБИД. Библиографическое описание аудиовизуальних материалов”. Цей стандарт набув чинності з 1 липня 2007 р. Він є базовим для системи стандартів, правил, методичних посібників зі складання бібліографічного опису.

Об’єктами опису є всі види опублікованих та неопублікованих документів на будь-яких носіях.

До складу бібліографічного опису входять такі області:

– область заголовка і відомостей про відповідальність;

– область видання;

– область специфічних відомостей;

– область вихідних даних;

– область фізичної характеристики;

– область серії;

– область примітки;

– область стандартного номера (або його альтернативи) й умов доступності.

Області опису складаються з елементів, які поділяють на обов’язкові та факультативні. В описі можуть бути лише обов’язкові елементи чи обов’язкові та факультативні.

Обов’язкові елементи містять бібліографічні відомості, що забезпечують ідентифікацію документа. Їх наводять у будь-якому описі. Якщо обов’язковий елемент є загальним для описів, що увійшли до бібліографічного посібника, і його винесено у заголовок бібліографічного посібника або його розділів, то він, як правило, не повторюється в кожному описі

(наприклад, ім'я автора в покажчику праць одного автора, ім'я видавця у видавничому каталозі, дата видання в хронологічному списку робіт і т. п.).

Факультативні елементи містять бібліографічні відомості, що дають додаткову інформацію про документ. Набір факультативних елементів визначає установа, яка упорядковує опис. Він має бути постійним для певного інформаційного блоку. Найбільш повно факультативні елементи наводять в описах для державних бібліографічних покажчиків, бібліотечних каталогів (у картковій та електронній формі), баз даних великих універсальних наукових бібліотек і центрів державної бібліографії. Факультативні елементи потрібні при підготовці літературного огляду чи оглядової статті – вони дають змогу орієнтуватися у великій кількості знайдених джерел інформації без потреби кожен раз звертатися до першоджерела.

Поля в бібліографічному описі відділяють знаками службової пунктуації (приписної пунктуації – рос. предписанной пунктуации):

. – крапка і тире;
. крапка;
, кома;
: двокрапка;
; крапка з комою;
… три крапки;
/ одинарний слеш;
// подвійний слеш;
( ) круглі дужки;
[ ] квадратні дужки;
+ знак "плюс";
= знак "дорівнює".

У кінці бібліографічного опису ставиться крапка.

Кожній області опису, окрім першої, передує крапка і тире, який ставиться перед першим елементом області. Якщо першого елемента немає, крапку і тире ставлять перед подальшим елементом, приписний знак якого в цьому випадку опускають. Виняток становлять круглі і квадратні дужки, які зберігаються і після знака області.

При повторенні області специфічних відомостей, області примітки й області міжнародного стандартного номера повто-

рюють і знак області – крапку і тире, а при повторенні області серії – дані про кожну серію закривають у круглі дужки без крапки і тире між ними.

Області опису можуть бути виділені різними шрифтами або записані з нового рядка. Тоді крапку і тире замінюють крапкою, яку ставлять у кінці попередньої області.

Для чіткішого розділення областей і елементів, а також для розрізнення приписної та граматичної пунктуації застосовують пропуски в один друкарський знак до і після знака, встановленого держстандартом. Виняток становлять крапка і кома – пропуски залишають лише після них.

Круглі та квадратні дужки розглядають як єдиний знак, пропуски ставлять перед відкриванням та після закривання дужок.

Якщо елемент не наводять в описі, то пропускають і приписаний йому знак. Елементи, граматично пов'язані в одному реченні, не розділяють приписною пунктуацією.

Частина елемента за потреби може бути пропущена. Пропуск частини елемента позначають знаком пропуску – трикрапкою з пропусками до і після неї. Відсутність області або елемента загалом трикрапкою не позначають.

Якщо сусідні елементи в межах однієї області мають бути наведені в квадратних дужках, то їх беруть у загальні квадратні дужки. Виняток становить загальне позначення матеріалу, яке завжди беруть в окремі квадратні дужки. Якщо суміжні елементи стосуються різних областей, то кожен елемент беруть в окремі квадратні дужки.

Усередині елементів зберігають пунктуацію, що не суперечить нормам мови, якою складено опис. Якщо елемент складається з декількох слів або фраз, що являють собою закінчені речення, їх наводять із тими знаками пунктуації, які вказано в документі. Якщо розділових знаків між словами або фразами немає, їх проставляють згідно з правилами наведення конкретних елементів опису, а також згідно з правилами граматики.

При поєднанні граматичного і приписного розділових знаків в описі наводять обидва знаки. Якщо елемент закінчується трикрапкою або крапкою в кінці скороченого слова, а приписна

пунктуація наступного елемента є крапкою або крапкою і тире, то крапку, що належить до приписної пунктуації наступного елемента, пропускають. Математичні, хімічні та інші знаки в описі відтворюють.

Бібліографічні відомості наводять у тому вигляді, в якому їх подано у джерелі інформації. Оскільки об'єктами опису є різні документи, визначення джерел інформації подано в узагальненому вигляді, тому що для кожного виду документів існує певна специфіка опису.

Загальне позначення матеріалу наводять в усталеній формі, вибираючи із поданого в національному стандарті ДСТУ ГОСТ 7. 1:2006 списку один із 14-ти термінів:

– відеозапис (videorecording);
– звукозапис (sound recording);
– образотворчий матеріал (graphic);
– карти (cartographic material);
– комплект (kit);
– кінофільм (motion picture);
– мікроформа (microform);
– мультимедіа (multimedia);
– ноти (music);
– предмет (object);
– рукопис (manuscript);
– текст (text);
– шрифт Брайля (braille);
– електронний ресурс (electronic resource).

З названих термінів вибирають один. Перевагу надають позначенню фізичної форми, у якій подано матеріал. Наприклад, якщо картографічний документ подано у вигляді мікроформи або електронного ресурсу, як загальне позначення матеріалу зазначають його форму.

Якщо серед декількох об'єктів на різних носіях не можна вибрати основний об'єкт, наводять позначення "[Мультимедіа]" чи "[Комплект]" (наприклад, інформація може бути розміщена на аудіокасеті, відеодиску та у вигляді короткого пояснювального тексту в брошурі).

Джерелом інформації для складання бібліографічного опису є документ загалом. У разі потреби можна навести відомості, запозичені поза документом, наприклад, з опублікованих бібліографічних записів про документ (каталоги бібліотек, бібліографічних покажчиків, баз даних) чи інших джерел поза документом (довідкових видань, авторитетних файлів, метаданих).

При складанні бібліографічного опису можна застосовувати скорочення слів і словосполучень, які мають відповідати вимогам ДСТУ 3582-97 "Скорочення слів в українській мові у бібліографічному описі. Загальні вимоги та правила", ГОСТ7. 12-93 "Библиографическая запись. Сокращения слов на русском язике. Общие требования и правила" та ДСТУ 7093:2009 "Система стандартів з інформації, бібліотечної та видавничої справи. Бібліографічний запис. Скорочення слів і словосполук, поданих іноземними європейськими мовами" (чинні в Україні як міждержавні). Заголовок опису формується згідно з ГОСТ 7.80-2000 "Бібліографічний запис. Заголовок. Загальні вимоги і правила складання". Варто зауважити, що за новими правилами в заголовку вказують тільки одного автора, котрий вніс найбільш суттєвий вклад у створення документа. Для хімії характерно, що авторами документів є зазвичай декілька осіб, отож їхні імена наводять у зоні відповідальності (слеш після назви).

6.1. Види бібліографічного опису

Залежно від об'єкта, на який потрібно скласти бібліографічний опис, вирізняють три види описів: монографічний, аналітичний, зведений. Монографічний складають на окремо виданий документ.

Приклад 6. *Монографічний бібліографічний опис:*

Педагогічна практика з хімії у середніх та вищих закладах: навч.-метод. посіб. [Текст] / В. І. Староста, В. М. Сомов, Ж. О. Кормош, О. В. Химинець. – Луцьк : РВВ "Вежа" Волин. держ. ун-ту ім. Лесі Українки, 2006. – 164 с.

Аналітичний бібліографічний опис складають для частини документа (статтю, главу, розділ) або на документ із Інтернету. Його укладають під узагальнювальною назвою, яку беруть за

основну. У відомостях, щодо назви, зазначають характеристику описаної групи статей чи матеріалів: статті, доповіді, відгуки, дискусії.

Приклад 7. *Аналітичний бібліографічний опис:*

Особенности анализа некоторых полупроводниковых материалов / [Ж. А .Кормош, Я. Р. Базель, Й. С. Балог, Я. И. Студеняк, О. Г. Воронич] // Second International Workshop "Relaxed, nonlinear and acoustic optical processes; Materials – Growth and optical properties". – RNAOPM'2005. June 01–05, 2005, Lutsk–Shatsk Lakes : Materials. – Lutsk: Volyn University Press "Vezha", 2005. – P. 92–94.

Зведений опис складають на багатотомне серіальне або продовжене видання.

Приклад 8. *Зведений бібліографічний опис:*

Кормош Ж. О. Електронна будова деяких п-диетиламіностирилових барвників, похідних 1,3,3-триметил-3Н-індолію [Текст] / Ж. О. Кормош, С. І. Корольчук, Я. Р. Базель // Наук. вісн. Ужгород. ун-ту. Серія : хімія. – 2002. – № 7. – С. 59–61

Залежно від повноти набору елементів бібліографічний опис буває: короткий, розширений, повний. Короткий – це опис, який складається тільки з обов'язкових елементів (основна назва, порядковий номер видання, місце та рік випуску видання, обсяг). Короткий опис частіше застосовують для бібліографічних посилань, які допомагають читачеві знайти об'єкт опису в бібліотеці; розширений – з обов'язкових та деяких факультативних елементів – для книжкових і статейних бібліографічних списків та покажчиків. Повний складається з обов'язкових та усіх факультативних елементів. Диференціація бібліографічного опису за повнотою набору елементів дає змогу більш раціонально і лаконічно застосувати його в різних сферах інформації.

Три основні функції – ідентифікувальна, інформаційна та пошукова – пов'язані між собою та використовуються в різних видах бібліографічного опису, залежно від ситуації пріоритети їхнього використання змінюються. Наприклад, при обробці

нових надходжень до бібліотеки чи інформаційного центру головну роль відіграє ідентифікувальна функція, а у видавничих каталогах – інформаційна.

Обліково-реєстраційна функція полягає в тому, що опис дає змогу здійснювати облік та реєстрацію документів і на загальнодержавному рівні, і на рівні тієї чи іншої установи. Організаційна функція – це формальні ознаки документа, які можуть бути основою для групування записів у конкретному пошуковому масиві.

Складаючи бібліографічний опис, потрібно дотримуватися певних вимог:

– точності – всі бібліографічні відомості мають відповідати інформації у документі;

– повноти даних – повний набір даних, що допомагають встановити відмінність документа від подібних;

– єдності опису – склад відомостей, форма і послідовність мають бути стабільними;

– стислості, оскільки бібліографічний опис є невеликим за обсягом;

– зрозумілості й чіткості.

Отже, бібліографічний опис – це записана за певними правилами множина бібліографічних даних, котрі ідентифікують документ.

Основною складовою частиною бібліографічного опису є бібліографічний запис.

Бібліографічний запис – це розгорнута бібліографічна характеристика видання (твору, документального джерела), у якій бібліографічний опис доповнено певними елементами: заголовком, анотацією чи рефератом, класифікаційними індексами, предметними рубриками, бібліотечними шифрами зберігання тощо. Саме такий вид бібліографічного опису варто використовувати для збору літератури.

6.2. Бібліографічний опис електронних документів

За змістом і характером інформації електронні дані поділяють на шрифтові, графічні, числові, звукові, текстові і демонстраційні; електронні програми – на прикладні, системні

та сервісні; до електронних даних і програм належать інтерактивні мультимедіа та онлайн-служби. За режимом доступу розрізняють ресурси локального доступу і видаленого доступу. Фізичними носіями електронного ресурсу є магнітні диски (жорсткі та гнучкі), оптичні диски, мікропроцесорні картриджі, касети і котушки з магнітною стрічкою. Інформаційний пошук – один із важливих додатків процесу комунікації. Основна мета пошуку – це знаходження повних текстів документів або відомостей про них.

Описи документів з електронних мереж складають за правилами формування аналітичних описів. У першій частині опису подають відомості про документ, у другій – його адресу. Наприклад, опис документа, взятого з Інтернету, має такий вигляд:

Приклад 9. *Бібліографічний опис документа з Інтернету:*

Влияние солнечной активности на работу ученых [Электронный ресурс]/ Л. Завальский : блог. – Режим доступа : http://www.inauka.ru/blogs/article40670. – 27.05.2010.

Різноманітність електронних видань, сайтів, інформаційних послуг, електронних інформаційних ресурсів в Інтернеті, їхня часта і швидка поява, а також їх швидке видалення дуже ускладнює завдання при описі та посиланні на них. Тому як у Росії, Україні, так і у всьому світі існує традиція: при посиланні на конкретний інформаційний ресурс Інтернету, вказувати дату його створення або зазначити адресу електронної пошти. Сьогодні день у зв'язку зі зростанням потреби у зазначенні посилань на вихідні дані джерела як для російськомовних видань, так і при оформленні публікацій для зарубіжних видань, актуальною є проблема бібліографічного опису документів, що належать до електронних ресурсів.

До повного бібліографічного опису статті входять такі елементи: прізвища й імена (і/або ініціали) авторів; назва статті; назва журналу; рік видання; номер тому; номер випуску; номери початкової й останньої сторінок, на яких опублікована стаття.

Приклад 10. *Повний бібліографічний опис журнальної статті:*

Киш П. П. Состояние некоторих цианиновых красителей в водном растворе / П. П. Киш, С. Г. Кременева // Журн. аналит. химии. – 1975. – Т. 30, №2. – С. 244–249.

У наукових публікаціях бібліографічний опис використовують тоді, коли потрібно дати посилання на роботу попередника. Згідно з "Вимогами до дипломних робіт, що захищаються на хімічному факультеті Волинського національного університету імені Лесі Українки", список використаних літературних джерел має оформлятися за формою, зазначеною вище.

Оскільки реально статтю можна виявити і за меншою кількістю параметрів, у наукових журналах зазвичай використовують скорочені бібліографічні описи – бібліограффічні посилання, причому кожен журнал може встановлювати власний зразок.

Бібліографічне посилання – сукупність бібліографічних даних про документ, що цитується, розглядається або згадується, необхідних і достатніх для його загальної характеристики, ідентифікації та пошуку.

До мінімального варіанта входить: прізвище й ініціали автора (всіх або тільки першого), скорочена назва журналу, номер тому, номер першої сторінки статті й рік видання, наприклад:

Приклад 11. *Скорочене посилання в журнальній статті:*

Stewart J. J. P., J. Comput. Chem. 10, 221 (1989)

Звертаємо увагу, що такий спрощений бібліографічний опис цілком придатний для виявлення статті в друкованому джерелі, для пошуку в електронній базі даних за прізвищем першого автора, за місцем публікації, але зовсім непридатний для тематичного пошуку.

Чинний держстандарт досить неоднозначний і описує здебільшого підходи до складання бібліографічних описів, не

даючи чіткого формату опису. Навіть у цьому стандарті наводяться рівнозначні варіанти опису одного й того ж видання.

У Росії введено ГОСТ Р 7.0.5–2008 “Библиографическая ссылка. Общие требования и правила составления”, котрий регламентує правила наведення бібліографічних посилань. Згідно з цим ГОСТом підготовка посилань на літературу є менш трудомісткою, ніж передбачено ДСТУ ГОСТ 7. 1:2006 України.

У більшості провідних наукових журналів, як правило, наводять у посиланнях всіх авторів публікації. Крім того, зазвичай пропускають назви статей, а назви журналів дають у скороченій формі.

6.3. Посилання на джерела інформації в тексті

Джерела інформації в списку літератури нумерують за порядком згадування в тексті роботи (в гуманітарних виданнях прийнято нумерувати позиції за алфавітом).

Посилання у тексті подають у квадратних дужках: [номер позиції] або [номер позиції бібліографії “кома” номер сторінки], або [номер позиції бібліографії “кома” номер сторінки “крапка з комою” номер позиції чергової бібліографії “кома” номер сторінки], або [від номера позиції “тире” по номер позиції], або [номер позиції “крапка з комою” номер позиції], або [номер позиції бібліографії “кома” номер сторінки “тире” номер сторінки].

Приклад 12. *Оформлення в тексті посилань на джерела інформації:*

[5] – посилання на джерело № 5;
[6, 12] – посилання на сторінку 12 джерела № 6;
[6, 12; 7, 15] – посилання на сторінку 12 джерела № 6 та сторінку 15 джерела № 7;
[7; 12] – посилання на джерела № 7 та № 12;
[8–13] – посилання на джерела з № 8 по № 13 включно (аналогічно запису [8; 9; 10; 11; 12; 13];
[9, 34–38] – посилання на сторінки з 34 по 38 джерела № 9.

6.4. Питання для самоперевірки

1. Які функції виконує бібліографічний опис?

2. Які знаки службової пунктуації використовують у бібліографічному описі?
3. Для яких видів джерел інформації використовують бібліографічний опис?
4. Які види матеріалів можна використовувати у бібліографічному описі?

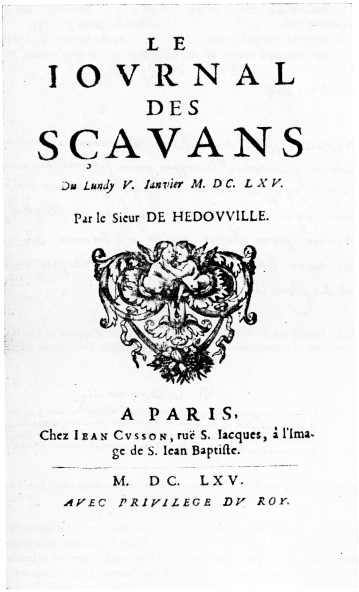
7. ІСТОРІЯ НАКОПИЧЕННЯ НАУКОВОЇ ІНФОРМАЦІЇ

Сучасна система наукових публікацій формувалася протягом століть і удосконалюється донині. На початковому етапі становлення науки результати досліджень підсумовували в рукописних книгах, доступних лише вузькому колу читачів. Книгодрукування сприяло тиражуванню наукових робіт і стимулювало їх стандартизацію. Суцільна нумерація сторінок, титульна сторінка, використання різних шрифтів для виділення смислових фрагментів, алфавітні покажчики, посилання на попередні роботи – все це ми сприймаємо як само по собі зрозумілі елементи книги, хоча вони формувалися і додавалися у процесі тривалої еволюції друкарських видань.

До середини XVII ст. у науці вже відбулися якісні зміни – експериментальний метод став нормою наукового дослідження, однак інформації щодо поставленого експерименту було порівняно небагато. Оскільки тоді у Європі вже існувала добре організована поштова система, тож не дивно, що колеги почали обмінюватися відомостями про результати досліджень за допомогою листів. (Відгомони тих часів відбилися в назвах деяких сучасних журналів, наприклад “Листи в ЖЕТФ” або “Letters in Organic Chemistry”).

На етапі, коли вчені почали об’єднуватися в наукові товариства, виникла потреба ознайомлювати ширшу аудиторію з матеріалами приватного наукового листування, що стало поштовхом для створення нового типу видання, яке б принципово відрізняло від звичайних на той час книг. Це були перші наукові журнали, тобто видання, що виходили періодично зі збереженням певного стандарту в їхній структурі.

5 січня 1665 р. побачив світ дебютний номер першого у світі наукового журналу – це був тижневик “Journal des scavans”. Орієнтовно через два місяці (6 березня 1665 р.) у Лондоні вийшов другий науковий журнал – “Philosophical Transactions of the Royal Society of London”, котрий видавався щомісяця. У цих журналах під однією обкладинкою друкували роботи різних авторів, не пов’язані між собою тематично і присвячені розгляду різних проблем із різних галузей науки. Протягом наступних 20-ти років на території тодішньої Європи з’явилося ще п’ять нових наукових журналів [6].


LE
IOVRNAL
DES
SÇAVANS
Du Lundy V. Ianvier M. DC. LXV.
Par le Sieur DE HEDOVVILLE.
A PARIS,
Chez IEAN CVSSON, ruë S. Iacques, à l'Image de S. Iean Baptiste.
M. DC. LXV.
AVEC PRIVILEGE DV ROY.

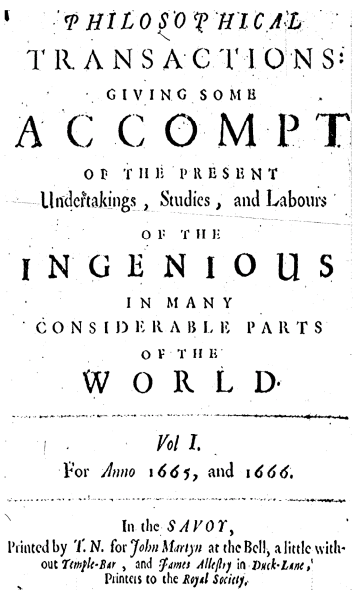

PHILOSOPHICAL
TRANSACTIONS:
GIVING SOME
ACCOMPT
OF THE PRESENT
Undertakings, Studies, and Labours
OF THE
INGENIOUS
IN MANY
CONSIDERABLE PARTS
OF THE
WORLD.
Vol I.
For *Anno* 1665, and 1666.
In the *SAVOY*,
Printed by *T. N.* for *John Martyn* at the Bell, a little without *Temple-Bar*, and *James Allestry* in *Duck-Lane*,
Printers to the *Royal Society*.

Рис. 10. *Титульні сторінки перших номерів наукових журналів – “Journal des scavans” (зліва) та “Philosophical Transactions” (справа)*

Наприкінці XVIII ст. відмічено першу кризу наукових видань – з’явилося багато журналів загальної тематики, і з цього часу почали створювати вузькоспеціалізовані наукові журнали (“The Neues medicinisches Wochenblatt für Aerzte” (1789 р.) –

перший спеціалізований журнал).

7.1. Зростання кількості публікацій від середини XVII до XX ст.

Отож, починаючи від середини XVII ст., кількість наукових статей та наукових журналів постійно зростала. Наприклад, 1732 р. голландський професор Герман Бургаве узагальнив і опублікував усю друковану спадщину з хімії у двох томах на 1434 сторінках. У 1825 р. шведський учений Берцеліус опублікував усі відомі доробки з хімії у восьми томах на 4150 сторінках. Сьогодні американський реферативний журнал “Chemical Abstracts”, що видається з 1907 р., публікує майже всю інформацію з хімії, при цьому перший мільйонний реферат був опублікований через 31 рік від дня заснування журналу, другий – ще через 18 років, третій – через наступні сім років, а четвертий – ще через чотири роки!

Приблизно такий же характер зростання кількості документів можна простежити і в інших сферах науки. Було зазначено, що зростання кількості документів має експоненціальний характер, при цьому щорічний приріст потоків науково-технічної інформації досягав 7–10 %. Кожні 10–15 років відбувалося подвоєння кількості науково-технічної інформації. Крива зростання кількості документів на цьому етапі можна описати експонентою. Така тенденція зберігалася приблизно до початку 90-х рр. XX ст. Потім зростання сповільнилося, і нині кількість наукових публікацій орієнтовно на 3 % на рік. У країнах Європи та Північної Америки зростання становить приблизно 1 %, а в країнах, що розвиваються, особливо в Ірані та Китаї, досягає 5 %. Причини цього явища буде висвітлено далі.

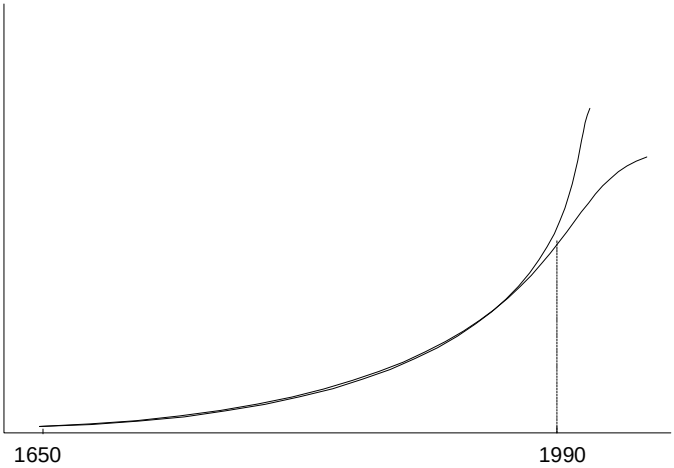


Рис. 11. *Зростання кількості публікацій середина XVII–XX ст.*

Збільшення кількості наукових публікацій значною мірою пов'язано зі зростанням загальної кількості вчених у світі. Протягом останніх п'яти століть спостерігається експоненціальне збільшення кількості науковців, воно триває і зараз. За даними ЮНЕСКО кількість учених у світі упродовж 2002–2007 р. зросла на 22 %, демонструючи щорічний приріст 4 %. При цьому темпи в країнах, що розвиваються, досягали 9 %, а в Китаї – аж 12 %. В Україні спостерігається зворотна тенденція – кількість учених протягом 1996–2007 рр. зменшилася на 47 %, демонструючи річне зменшення на 5,5 %!

Можна передбачити, що експоненційне зростання не може тривати безкінечно. Настане час, коли він сповільниться, а потім – і зупиниться. За даними ЮНЕСКО (за 2007 р.), частка вчених становить у середньому 0,1 % загальної кількості населення, в розвинутих країнах – 0,2–0,5 % (в Японії – 0,55 %, в Україні – 0,17 % і далі знижується). Кількість вчених у розвинутих країнах стабілізується десь на рівні 0,5 %, отже перспективи інтенсивного розвитку науки ще далеко не вичерпані.

Як уже зазначалося, на сьогодні темпи зростання кількості наукових публікацій усе ще високі, тому це викликає у споживачів інформації тривогу: як зберігати й обробляти документи, як знайти серед них потрібний?

Становище може здатися безвихідним: поки діє у сфері документування закон експоненціального зростання їхньої кількості, він різко загострює і проблему зберігання інформації, і проблему її поширення.

Проте в процесах зростання і розповсюдження інформації спостерігаються певні особливості.

Наприкінці 40-х рр. XX ст. Дж. Ципф (G. K. Zipf), зібравши величезний статистичний матеріал, показав, що розподіл слів природної мови підкоряється одному простому законові, який можна сформулювати так: якщо до якого-небудь достатньо великого тексту скласти список усіх слів, що вживаються в ньому, потім розмістити ці слова в порядку зменшення частоти їх використання й пронумерувати від 1 (порядковий номер слова, що найбільш часто вживається) до останнього, то для будь-якого слова добуток його порядкового номера (рангу) в такому списку і частоти, з якою його використовують у тексті, буде величиною постійною. Вона буде мати однакове значення для будь-якого слова із цього списку. Аналітично закон Ципфа можна виразити у вигляді:

$$f \cdot r = c, \qquad (1)$$

де f – частота вживання слова в тексті; r – ранг (порядковий номер) слова у списку; c – емпірична постійна величина.

Отриману залежність графічно виражають гіперболою. Дослідивши таким способом найрізноманітніші тексти і мови, зокрема мови тисячолітньої давності, Дж. Ципф побудував такі залежності для кожної з них, при цьому всі криві мали однакову форму – форму “гіперболічних сходів”, тобто при заміні одного тексту іншим, загальний характер розподілу не змінювався.

Закон Ципфа відкрито експериментально, однак пізніше Бенуа Мандельброт (Benoît Mandelbrot) запропонував його теоретичне обґрунтування. Він вважав, що можна порівнювати письмову мову з кодуванням, причому всі знаки повинні мати певну “вартість”. Виходячи з вимог мінімальної вартості повідомлень, Б. Мандельброт математичним способом створив аналогічну із законом Ципфа залежність:

$$f \cdot r^{\gamma} = c, \qquad (2)$$

де γ – величина (близька до одиниці), яка може змінюватися

залежно від властивостей тексту.

Дж. Ципф та інші дослідники встановили, що такій залежності підпорядковується не тільки всі природні мови світу, а й інші явища соціального та біологічного характеру: розподіл вчених за кількістю опублікованих ними статей (А. Лотка, 1926 р.), міст США за кількістю населення (Дж. Ципф, 1949 р.), населення за розміром доходів у капіталістичних країнах (У. Парето, 1897 р.), біологічних родів за чисельністю видів (Дж. Уїлліс, 1922 р.) та ін.

Важливим є той факт, що й документи з будь-якої галузі знань поділяють згідно з цим законом; окремим його випадком є закон Бредфорда, безпосередньо пов'язаний уже не з розподілом слів у тексті, а із розподілом документів усередині певної тематичної ділянки.

Англійський хімік та бібліограф С. Бредфорд, досліджуючи статті з прикладної геофізики та мастил, зазначив, що розподіли наукових журналів, які містять статті про мастила і журналів, що містять статті з прикладної геофізики, мають однаковий вигляд. На підставі з'ясованого факту С. Бредфорд сформулював закономірність розподілу публікацій за виданнями.

Основний сенс закономірності полягає в тому що, якщо наукові журнали розмістити в послідовності зменшення кількості статей з конкретного питання, то журнали в отриманому списку можна розбити на три зони так, щоб кількість статей з певної ділянки наук у кожній із них була однаковою. При цьому в першу зону, так звану зону ядра, входять профільні журнали, власне присвячені певній тематиці (кількість профільних журналів у зоні ядра велика). Другу зону утворюють журнали, частково присвячені окремій галузі, причому їхня кількість істотно зростає порівно з кількістю журналів у ядрі. Третя зона, найбільша за кількістю видань, об'єднує журнали, тематика яких досить далека від обраного предмета.

Так, при однаковій кількості публікацій з певної тематики в кожній зоні кількість найменувань журналів різко зростає при переході від однієї зони до іншої. С. Бредфорд установив, що кількість журналів у різних зонах вираховують за степеневим законом:

$$N_1 : N_2 : N_3 = 1 : a : a^2, \qquad (3)$$

де N_1 – кількість журналів у першій зоні, N_2 – кількість журналів у другій зоні, N_3 – кількість журналів у третій зоні, а – константа, залежна від тематики.

Цю залежність називають законом Бредфорда.

Б. Вікері уточнив модель С. Бредфорда. Він установив, що журнали, проранжировані у послідовності зменшення у них статей з конкретного питання, можна розбити не на три зони, а на будь-яку потрібну кількість зон. Якщо періодичні видання розмістити у послідовності зменшення в них кількості статей із конкретного питання, то в отриманому списку можна виділити декілька зон, кожна з яких містить однакову кількість статей:

$$N_{1x} : N_{2x} : N_{3x} : N_{4x} : \ldots = 1 : a : a^2 : a^3 : a^4 : \ldots, \qquad (4)$$

де N_{1x} – кількість журналів, які містять x статей, N_{2x} – кількість журналів, що містять $2x$ статей, N_{3x} – кількість журналів, де вміщено $3x$ статті, N_{4x} – кількість журналів, що містить $4x$ статті, a – константа.

Цей вираз називають ще законом Бредфорда в тлумаченні Б. Вікері.

Якщо закон Ципфа характеризує багато явищ соціального й біологічного характеру, то закон Бредфорда – це специфічний випадок розподілу закону Ципфа для системи наукових періодичних видань.

З цих закономірностей можна зробити важливі практичні висновки: якщо розмістити будь-які періодичні видання у послідовності зменшення кількості статей з певним профілем, то, згідно з Бредфордом, їх можна поділити на три групи, що містять однакову кількість статей. Нехай ми відібрали групу із семи найменувань журналів, що займають перші 7 місць в отриманому списку, тоді для того, щоб подвоїти кількість статей за профілем, який нас цікавить, доведеться додати до тих, що є, ще $7a$ найменувань журналів. Якщо $a = 5$, то кількість найменувань дорівнюватиме 35, тоді загальна кількість найменувань періодичних видань складатиме 42, що, звичайно, значно більше, ніж 7. При спробі ж отримати втричі більшу кількість статей нам доведеться охопити вже $7 + 7 \cdot 5 + 7 \cdot 5^2 = 217$ найменувань! З них третину

статей, що нас цікавлять, зосереджено всього у семи журналах, тобто статті розподіляються нерівномірно за найменуваннями журналів. З одного боку, спостерігається накопичення значної кількості статей із певної тематики в декількох профільних журналах, з іншого – їх розпорошення у величезній кількості видань із суміжної або далекої від обраної тематики, тоді як на практиці потрібно виявити основні джерела в галузі, що нас цікавить, а не випадкові видання.

Закономірності накопичення і розпорошення науково-технічної інформації у сфері документування дають змогу вибирати саме ті видання, які найбільш ймовірно містять публікації, що належать до певної галузі знань. У масовому процесі інформаційного забезпечення в масштабах країни використання цих закономірностей дає змогу скоротити величезні витрати державного бюджету.

Таке розпорошення публікацій не можна оцінювати лише як шкідливе явище, адже в таких умовах поліпшуються можливості для міжгалузевого обміну інформацією.

Спробу сконцентрувати всі публікації з одного профілю у декількох журналах, тобто не допустити їх розпорошення, матиме негативні наслідки, не кажучи вже про те, що достовірно визначити приналежність документа до того або іншого профілю не завжди можливо.

Величезна заслуга Дж. Ципфа і С. Бредфорда полягає в тому, що вони започаткували дослідження документальних інформаційних потоків, які є сукупністю наукових документів – публікацій та неопублікованих матеріалів. Подальші дослідження показали, що можна визначати не тільки кількісні параметри сукупностей наукових документів, а й сукупностей елементів ознак наукових документів: авторів, термінів, індексів класифікаційних ознак, найменувань видань, тобто найменувань елементів, що характеризують зміст наукових документів. Наприклад, можна розмістити журнали в послідовності зменшення кількості авторів, що друкуються в них, у послідовності зменшення середнього обсягу статей, що публікуються, або упорядкувати сукупність документів за будь-яким елементом.

Упорядкованість задається ранжируванням (порядком розміщення) найменувань елементів за частотою їх появи у послі-

довності її зменшення. Така впорядкована сукупність найменувань елементів називається ранговим розподілом. Розподіли, які свого часу вивчав Ципф, – це типові приклади рангових розподілів. Виявилося, що вид рангового розподілу і його будова характеризують ту сукупність документів, до якої належить цей ранговий розподіл. З'ясувалося, що найчастіше рангові розподіли мають вигляд закону Ципфа з поправкою Мандельброта (формула 2, стор. 53).

Уточнили, що коефіцієнт γ – величина змінна. Постійність коефіцієнта γ зберігається тільки на середній ділянці графіка розподілу. Ця ділянка набуває форми прямої, якщо графік побудувати в логарифмічних координатах. Ділянка розподілу з $\gamma = \text{const}$ називається центральною зоною рангового розподілу.

Який же сенс закладено в існуванні трьох явно відмінних зон рангових розподілів? Якщо розподіл належить до термінів, що є складовими частинами певної ділянки знань, то центральна зона рангового розподілу містить найбільш загальновживані і найбільш характерні для певної галузі знань терміни, які в сукупності засвідчують її специфічність, відмінність від інших наук, охоплюють її основний зміст. В інших зонах зосереджено терміни, що порівняно рідко вживають у певній ділянці наук.

Отож, основу лексики певної галузі знань зосереджено в центральній зоні рангового розподілу. За допомогою термінів ядерної зони вона поєднується з більш загальними її частинами, а інші зони містять зв'язки з іншими галузями науки.

7.2. Вплив сонячної активності на роботу вчених

За останні сто років було багато написано про зв'язок між подіями на Сонці й на Землі. Періодичність та щільність річних кілець дерев, кліматичні коливання, інтенсивність розмноження людей і тварин, міграційні потоки, економічні й політичні цикли, війни й революційні суспільні перетворення – ці та багато інших біологічних і суспільних явищ так чи інакше пов'язані зі змінами в енергії Сонця. Прекрасною ілюстрацією впливу Сонця на протікання всіх живих процесів, що відбуваються на Землі, можуть послугувати висновки професора Сіднея Чепмена, підкріплені математично точними кількісними вимірюваннями. Аргументація вченого завдає сильного удару

по позиціях скептиків, що сумніваються в циклічній дії Сонця на біосферу Землі. Відомий англійський вчений С. Чепмен (Sydney Chapman), відзначаючи у 1968 р. своє вісімдесятиліття, підрахував, що за 58 років своєї наукової діяльності він опублікував 450 робіт з фізики верхніх шарів атмосфери Сонця. Оскільки продуктивна діяльність цього видатного ученого тривала протягом більш ніж п'яти циклів сонячної активності, він зробив спробу порівняти частоту сонячних циклів із коливаннями власної наукової діяльності. Результати порівняння зображено у вигляді графіка (рис. 12).

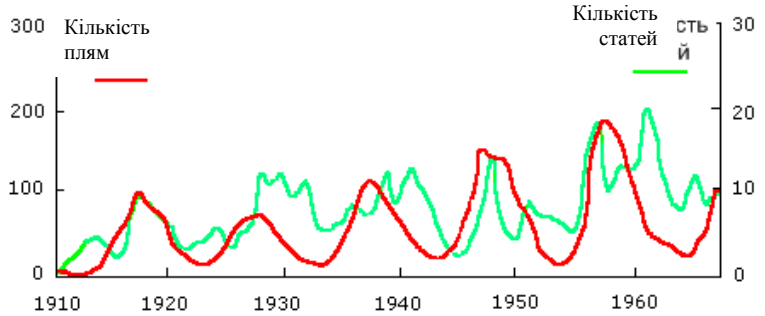


Рис. 12. *Залежність кількості наукових статей професора Чепмена від кількості плям на Сонці*

З аналізу графіків наукової діяльності професора Чепмена можна зробити щонайменше три важливі висновки:

1) кількість публікацій збільшується і зменшується відповідно до кількості плям на сонці з 11-річним періодом;

2) деяке запізнення, приблизно на рік, графіка публікацій ученого відповідно до сонячної активності можна пояснити роботою над рукописом і редакційною тяганиною, що вимагає значних витрат часу (у нашій країні час затримки зріс би в десятки разів);

3) висота двох суміжних графічних максимумів поступово зростає, що засвідчує збільшення й енергії Сонця, і професора Чепмена. Очевидно також, що з віком наукова продуктивність ученого тільки зростала – отже старість зовсім не стає

перешкодою для розумової діяльності.

Додамо, що метод професора Чепмена можна успішно використовувати для статистичної обробки багатьох інших закономірностей. Наприклад, для вивчення впливу ДТП (дорожньо-транспортних пригод) на роботу ДПС (дорожньо-патрульної служби), впливи злочинності на прокуратуру, стихійних лих на МНС тощо (за матеріалами статті W. H. Campbell "Transaction of Geophysical Union", 1968, v. 49, '4, pp. 609–610 [10])

7.3. Бібліотечна криза

Інший наслідок зростання кількості публікацій – бібліотечна криза. Унаслідок зростання кількості журналів та збільшення підписних цін університетські бібліотеки більше не в змозі здійснювати підписку на всі потрібні журнали. Зменшення попиту спричиняє загальне підвищення цін на наукові журнали. Отож одержуємо замкнуту спіраль, яка призводить до того, що навіть у найбільшій бібліотеці не завжди є усі найвідоміші наукові журнали.

7.4. Комерціалізація інформації з хімії

Оскільки доступ до більшої частини інформації з хімії в Інтернеті платний, це обмежує її поширення. При цьому її автори – науковці – до цього зазвичай, не причетні. Більше того, вони якраз і зацікавлені у вільному розповсюдженні інформації та її обміні, якщо, звичайно, йдеться про патентоспроможні рішення. А бар'єром на шляху вільного розповсюдження наукової інформації стають ті ж наукові журнали. І якщо раніше, до появи Інтернету й інших мережевих ресурсів, ці журнали для ученого були чи не єдиним способом сповістити наукову громадськість про результати своїх досліджень, то сьогодні журнали, які друкуються на папері, стають дедалі більшим анахронізмом. Тільки завдяки величезній інерції серед науковців такі журнали не втрачають своєї актуальності. Тоді як спосіб поширення і зберігання друкованої інформації вкрай неефективний, трудомісткий і дорого коштує, електронні засоби, передусім Інтернет, дають змогу виконувати ці завдання набагато швидше, простіше й дешевше. Хоча багато наукових журналів, прагнучи не відстати від технічного прогресу,

дублюють свої паперові випуски в електронному вигляді, проте обмежуючи доступ до них, адже їм треба окупити витрати на паперові видання. Отож маємо явну суперечність між інтересами творців та споживачів наукової інформації, з одного боку і, як посередників між ними, наукових журналів – з іншого.

Паперові журнали, очевидно, відживають свій вік і в не дуже далекому майбутньому, найімовірніше, зникнуть зовсім. Якщо художня література ще зможе існувати в друкованому вигляді і звичка посидіти з книжкою не відімре так швидко, то наукова інформація цілком може обійтися без паперових носіїв. Питання в тому, як протидіяти прагненню посередників спекулювати інформацією, яка створена не ними і не для них.

Вільний доступ до статті означає, що з нею ознайомиться більше людей, а отже буде більше шансів, що її процитують. Коли публікується якийсь матеріал у науковому журналі, передається право на одноразове його використання, при цьому, як правило, основні авторські права залишаються за авторами (якщо немає окремої угоди про повну передачу авторських прав журналу). Тому не варто заспокоюватися на тому, що статтю, опубліковано в журналі, паралельно потрібно викласти її в Інтернеті. Зараз усі серйозні наукові установи, ВНЗ і навіть окремі кафедри та лабораторії мають власні веб-сайти в Інтернеті. Інша справа, що багато з них мають досить примітивний зміст унаслідок того, що працівники цих організацій не заповнюють їх матеріалами. Якщо ж розмістити там свою статтю з якихось причин неможливо, то варто створити власний веб-сайт або домашню сторінку. Це можна зробити дуже просто й безкоштовно. Виклавши статтю в Інтернеті, можна одержати низку переваг: по-перше, її буде проіндексовано головними пошуковими системами типу Google не тільки за назвою, авторам або ключовим словам, а й за текстом статті, тобто будь-яке слово може стати ключовим при пошукові. Отже, ймовірність того, що про ваше дослідження дізнається багато людей, різко зростає порівняно зі звичайною публікацією в журналі з невеликим тиражем і обмеженим доступом до електронної версії. По-друге, навіть багато з тих, хто знатиме про статтю в журналі, не зможе її прочитати, якщо не є передплатником цього видання (а передплатників наукових

журналів зазвичай небагато), у разі ж вільного доступу в Інтернеті її прочитають усі, хто в цьому зацікавлений. По-третє, викладаючи свої публікації для вільного доступу, варто чекати аналогічної поведінки від своїх колег, що, цілком можливо, викличе ланцюгову реакцію вільного поширення інформації.

Але що робити, якщо статті у вільному доступі немає? Тут варто звернутися безпосередньо до авторів. Раніше, до появи Інтернету, існувала практика обміну інформацією: можна було написати авторові статті звичайного листа й він, як правило, надсилав копію статті. Навіть журнали з обмеженим доступом свій зміст публікують відкрито і в багатьох з них указують контактну інформацію авторів. Інтернет та електронна пошта і тут істотно полегшують завдання: можна написати електронного листа авторові з проханням надіслати електронну версію його публікації. Зробити це зараз простіше й дешевше, ніж раніше надсилати копію, і практика показує, що переважна більшість авторів відгукується на такі прохання. Якщо ж у журналі немає контактної інформації авторів, то знову-таки за допомогою пошукових систем в Інтернеті майже завжди можна знайти електронну адресу автора або установи, де він працює.

Революційним почином можна вважати недавнє створення сервера "Chemistry Preprint Server" (http://www.sciencedirect.com/preprintarchive), що є одночасно веб-архівом і способом поширення наукових статей у галузі хімії (щось подібне до депонування рукописів). Цей сервер створено в межах проекту Chemweb (http://www.chemweb.com), і всі зареєстровані користувачі можуть викладати на сервері свої статті й користуватися всім архівом. Можна викладати і результати закінчених досліджень, і проміжні результати роботи. Для користування ресурсом потрібно безкоштовно зареєструватися за адресою https://www.sciencedirect.com/preprintarchive/activate/members. Матеріали, що зберігаються в архіві "Chemistry Preprint Server", індексує низка пошукових систем, зокрема Scirus.

7.5. Питання для самоперевірки

1. Що стало причиною швидкого прогресу науки?
2. Які причини зростання кількості наукових публікацій?

3. У чому суть закону Ципфа?
4. У чому суть закону Бредфорда?
5. Що пов'язує закони Ципфа та Бредфорда?

8. ОЦІНЮВАННЯ ЕФЕКТИВНОСТІ НАУКОВИХ ПУБЛІКАЦІЙ

Унаслідок зростання кількості публікацій постала проблема оцінювання якості наукової роботи. Для цього запропоновано використовувати різноманітні індекси цитувань.

8.1. Індекс цитувань

Індекс цитувань – ключовий показник, що широко використовується в усьому світі для оцінки роботи дослідників і наукових колективів. Він оцінює вплив науковця або організації на світову науку, визначає якість наукових досліджень.

Індекс цитування наукових статей (ІЦ) – показник, який визначає значущість певної статті й обраховується на основі подальших публікацій, що спираються на цю роботу.

Перший індекс цитування пов'язаний з юридичними посиланнями і датується 1873 р. (Shepard's Citations). У 1960 р. Інститут наукової інформації (ISI), що його заснував Євген Гарфілд (Eugene Garfield), увів перший індекс цитування для статей, опублікованих у наукових журналах, поклавши початок такому ІЦ, як "Science Citation Index" (SCI) (англ), і потім включивши до нього індекси цитування із суспільних наук ("Social Sciences Citation Index", SSCI) і мистецтв ("Arts and Humanities Citation Index", AHCI). Починаючи з 2006 р., з'явилися й інші джерела подібних даних, наприклад Google Scholar. SCI повністю представлено в онлайн-проекті Web of Science (англ).

Статистичні дані щодо індексу цитувань відомих учених світу розміщено у "Science Citation Index" (SCI) та "Journal Citation Reports" (JCR), що випускає Institute for Scientific Information (ISI), Філадельфія, США.

8.1.1. Різновиди

Індекс цитування (Science Citation Index, – SCI) – система Філадельфійського інституту наукової інформації, за основу якої взято зв'язки між документами за прямими, зворотніми і перехресними посиланнями (цитуванням).

Система для обчислення SCI (або його інтернет-версія ISI Web of Knowledge) містить бібліографічні описи всіх статей із наукових журналів, що входять до переліку JCR, і висвітлює переважно публікації із фундаментальних галузей науки в провідних міжнародних і національних журналах.

JCR ("Journal Citation Reports") – бібліометричний довідник, у якому подають повну й різноманітну статистику цитування наукових журналів, включно із широкий спектром показників використання журналів ученими різних країн.

8.2. Імпакт-фактор

Імпакт-фактор (ІФ, або IF) – числовий показник важливості наукового журналу. З 1960-х рр. його щорічно розраховує Інститут наукової інформації (Institute for Scientific Information (ISI), придбаний у 1992 р. корпорацією "Thomson", і нині має назву "Thomson Scientific"), і публікується в журналі "Journal Citation Report". Згідно з ІФ (переважно в інших країнах, але останнім часом усе більше і в Україні) оцінюють рівень журналів, якість статей, у них опублікованих, дають фінансову підтримку дослідникам, розширюючи штат співробітників. Імпакт-фактор має хоча й великий, але неоднозначно сприйнятний вплив на оцінювання результатів наукових досліджень.

8.2.1. Методика розрахунку

Розрахунок імпакт-фактору засновано на трирічному періоді, наприклад, імпакт-фактор журналу в 2008 р. обчислюватиметься так:

$$I_{2008} = \frac{A}{B}, \qquad (5)$$

де:

A – кількість цитувань протягом 2008 р. в журналах, що відстежує Інститут наукової інформації;

B – кількість статей, опублікованих у цьому журналі в 2006–2007 рр.

У методиці розрахунку є декілька нюансів: Інститут наукової інформації при розрахунках не бере до уваги деякі типи статей (повідомлення, листи, списки друкарських помилок тощо), і для нових журналів у такому разі імпакт-фактор іноді розраховують тільки на дворічний період.

8.2.2. Переваги та недоліки

ІФ журналу залежить від галузі досліджень і його типу; з роками він може помітно змінюватися, наприклад, опускаючись до гранично низьких значень при зміні назви журналу, та ін. Проте на сьогодні ІФ є одним з важливих критеріїв, за яким можна зіставляти рівень наукових досліджень у суміжних галузях знань. Наприклад, інвестор наукового дослідження може виявити бажання порівняти результати дослідників для оцінки перспектив своїх внесків, для цього й використовують об'єктивні числові показники, такі як імпакт-фактор – тому на подібні вимірювання й існує попит.

Позитивні якості імпакт-фактору:

– значне охоплення наукової літератури – індексують понад 8400 журналів із 60 країн;

– результати публічні та досяжні;

– простота в розумінні й використанні;

– журнали з високим ІФ, як правило, мають жорсткішу систему рецензування, порівняно з виданнями з низьким ІФ.

Водночас імпакт-фактор не ідеальний. Наприклад, не зрозуміло, наскільки достовірно кількість цитувань засвідчує про якість статті. Крім того, в журналах з тривалим періодом публікації виявляють статті, які спираються на публікації, що не потрапляють до трирічного інтервалу. Дійсно, в деяких журналах проміжок часу між ухваленням статті й публікацією становить понад два роки, залишаючи лише рік на посилання, які враховують при розрахунках. З іншого боку, збільшення тимчасового проміжку, в якому беруть до уваги цитування, зробить імпакт-фактор менш чутливим до змін.

Найбільш очевидні недоліки імпакт-фактору такі:

– кількість цитувань насправді не відображає якості дослідження, втім, як і кількість публікацій;

– проміжок часу, протягом якого враховують цитування, дуже короткий (класичні статті часто цитують навіть через декілька десятиліть після публікації);

– природа результатів у різних ділянках дослідження приводить до різної періодичності публікації результатів, які впливають на імпакт-фактори. Так, наприклад, медичні журнали зазвичай мають вищі імпакт-фактори, ніж математичні.

8.3. Індекс Хірша

Індекс Хірша, або h-індекс – наукометричний показник, який запропонував у 2005 р. американський фізик Джорж Хірш (Jorge E. Hirsch) з університету Сан-Дієго (Каліфорнія). Індекс Хірша є кількісною характеристикою продуктивності вченого, заснованою на кількості його публікацій та періодичності цитувань.

8.3.1. Визначення

Індекс обчислюється на основі розподілу цитувань робіт певного дослідника. Учений має індекс h, якщо h з його Np статей цитують як мінімум h разів кожну, тоді як інші (Np – h) статті цитують не більш ніж h разів кожну.

Іншими словами, учений з індексом h опублікував h статей, на кожну з яких дали посилання як мінімум h разів. Отож, якщо у цього дослідника опубліковано 100 статей, на кожну з яких є лише одне посилання, його h-індекс дорівнює 1. Таким же буде h-індекс дослідника, що опублікував одну статтю, на яку послалися 100 разів. Водночас (реалістичніший випадок), якщо у дослідника є одна стаття з дев’ятьма цитуваннями, дві статті з вісьмома цитуваннями, три статті з сімома цитуваннями, девять статей з одним цитуванням кожної з них, то його h-індекс дорівнює 6. Звичайний розподіл кількості публікації N(q) залежно від кількості їх цитувань q дуже приблизно зображено за допомогою гіперболи:

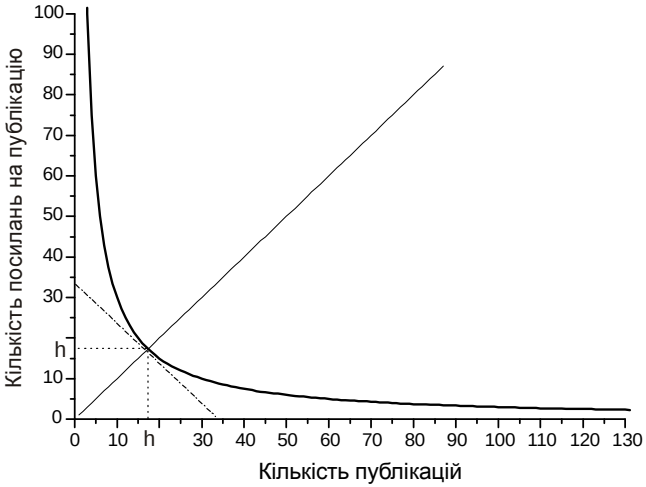


Рис. 13. *Визначення індексу Хірша за розподілом публікацій за кількістю посилань*

Координата точки перетину цієї кривої з медіаною $N(q) = q$ і дорівнюватиме індексові Хірша.

Індекс Хірша розроблено для того, щоб отримати більш адекватну оцінку наукової продуктивності дослідника, ніж можуть дати такі прості характеристики, як загальна кількість публікацій або загальна кількість цитувань. Індекс добре працює лише тоді, коли порівнюють дослідників, що працюють в одній галузі, оскільки традиції, пов'язані з цитуванням, відрізняються в окремих галузях науки (наприклад, у біології та медицині h-індекс набагато вищий, ніж у фізиці). У нормі h-індекс фізика приблизно дорівнює тривалості його наукової кар'єри в роках, тоді як у видатного фізика він удвічі вищий. Хірш вважає, що у фізиці (для реалій США) h-індекс, що дорівнює 10–12, може слугувати одним із визначальних чинників для вирішення питання про надання дослідникові постійної позиції у великому дослідницькому університеті; рівень дослідника з h-індексом, який сягає 15–20, достатній для членства в Американському фізичному товаристві; індекс 45 і більше може означати членство у Національній академії наук США. На думку Хірша, успішний учений-фізик має h-індекс понад 10 (h-індекс самого

Хірша дорівнює 49).

Як і будь-який формальний показник, h-індекс має свої переваги та недоліки. До його переваг належить той факт, що він буде однаково низьким як і автора однієї надпопулярної статті, і для автора безлічі робіт, процитованих не більше одного разу. Індекс Хірша дає змогу відсівати і так званих "випадкових співавторів"; цей показник буде високим лише для тих, у кого достатньо публікацій, і всі вони (або принаймні багато з них) достатньо затребувані, їх часто цитують інші дослідники.

Індекс Хірша обчислюють, використовуючи загальнодоступні (Google Scholar) або платні (Scopus) бази даних Інтернету. Хоча помічено, що інколи при розрахунку різні автори сприймаються як один (Google Scholar), або один і той самий автор потрапляє в різні категорії (Scopus), відповідно статті розпорошуючи, і, як наслідок, отримують некоректний h-індекс.

Зрозуміло, що індекс Хірша недосконалий. Неважко змоделювати ситуацію, коли h-індекс дає абсолютно неправильну оцінку значущості дослідника. Зокрема, коротка кар'єра ученого призводить до недооцінки значущості його робіт. Якби Альберт Ейнштейн помер на початку 1906 р., його h-індекс зупинився б на 4 або 5, незважаючи надзвичайно високу значущість статей, опублікованих ним у 1905 р.

Більший індекс Хірша, який вказує на значущість вченого для розвитку світової науки, мають науковці, які продуктивно працюють протягом тривалого часу.

8.4. Питання для самоперевірки

1. Що таке індекс цитувань?
2. У чому полягає важливість цитування наукових робіт?
3. Яка роль цитатування у розвитку науки?
4. Назвіть слабкі та сильні сторони індексу цитувань.
5. Дайте визначення індексу Хірша
6. Які переваги та недоліки індексу Хірша?

9. ІНТЕРНЕТ

Інтерне́т (від англ. Internet, дослівно – міжмережа) – всесвітня

система об'єднаних комп'ютерних мереж, побудована на використанні протоколу IP і маршрутизації пакетів даних.

Інтернет утворює глобальний інформаційний простір, слугує фізичною основою доступу до веб-сайтів і багатьох систем (протоколів) передачі даних.

Сьогодні, вживаючи слово "Інтернет", найчастіше мають на увазі саме мережу веб-сторінок і доступну завдяки їй інформацію, а не власне фізичну мережу.

9.1. Історія Інтернету

Перша комп'ютерна мережа з'явилася 1973 р. у США і отримала назву "ARPAnet" (Advanced Research Project Agency – Управління передових досліджень Міністерства оборони США) на ім'я організацій, якій Міністерство оборони США доручило її створення.

Потім у США з'являються інші обчислювальні мережі, наприклад NASA, DOE та ін. У 1983 р. з ARPAnet виокремили мережу Міністерства оборони США MILNET, що й далі взаємодіяла з мережею, котра її породила. У результаті вперше з'являється термін "Інтернет", що спочатку позначав спілкування між собою двох мереж MILNET і ARPAnet.

У 1983 р. Пол Мокапетріс (Paul Mockapetris) запропонував систему розподілених доменних імен (англ. Domain Name System, DNS), яку впроваджено протягом року (детальніше на с. 73). Через 22 роки, у травні 2005 р., він одержав за це відкриття пожиттєву премію від компанії ACM Sigcomm.

У 1984 р. до ARPAnet приєднали Національний науковий фонд США (National Science Foundation, NSF), а в 1986 р. на основі використання каналів телефонного зв'язку створено мережу NFSNET (National Science Foundation Network), що об'єднала учених із різних університетів США.

У 1987 р. ця мережа перетворилася в NFSNET Backbone (базова частина), що об'єднала цифровими швидкодійними каналами (1,45 Mbps – Мбіт/с) 13 територіальних центрів.

У 1990 р. Тімоті Джон Бернерс-Лі (англ. Timothy John Tim Berners-Lee), котрий працював тоді у Європейській раді з

ядерних досліджень (ЦЕРН), запропонував систему URL для ідентифікації ресурсів мережі. У тому ж році проект, заснований на ідеї взаємодії незалежних мереж, почали називати Інтернетом (англ. Internet), тобто "мережею мереж" – структурою, що об'єднує інші (опорні) обчислювальні й інформаційні комп'ютерні мережі.

У 1991 р. ARPANET припиняє своє існування, а створена в її межах мережа Інтернет існує і об'єднує безліч мереж у всьому світі. 6 серпня 1991 р з'являється перший у світі веб-сайт (http://info.cern.ch), який створив той же Т. Дж. Бернерс-Лі. Цю дату вважають датою запровадження мережі Інтернет.

Практично будь-яка комп'ютерна мережа, що забезпечує взаємодію з іншими комп'ютерними мережами у світі і працює за протоколом TCP/IP, є частиною (фрагментом або сегментом) Інтернету. У Росії першою організацією, що отримала на початку 1980-х рр. доступ до Інтернету, стає Інститут атомної енергії ім. І. В. Курчатова (IAE). У 1989 р. для розробки програмного забезпечення і побудови локальних комп'ютерних мереж створюють компанію ДЕМОС. У 1990 р. починає функціонувати мережа користувачів UNIX, що отримала назву Релком. Вона забезпечила зв'язок між IAE і ДЕМОСОМ, а також підключення до європейської мережі користувачів UNIX–EUnet. Окрім Relcom, у наший країні успішно існували й працювали декілька регіональних комп'ютерних мереж, наприклад, Sprint, Glasnet, Rosnet та інші.

До кінця XX ст. можливостей Інтернету вже не вистачало. У 1996 р. декілька десятків американських університетів виступило з ініціативою створення системи "Internet 2", призначеної для розробки й поширення передових мережевих технологій серед наукових та освітніх структур, зокрема методів організації високоякісних цифрових телепередач, дистанційного керування науковим устаткуванням (телескопами, мікроскопами й ін.), віртуальної реальності і систем розподілених розрахунків.

9.2. Протоколи та адреси

Щоб різні комп'ютери могли разом працювати та здійснювали обмін інформацією, вони мають "розмовляти однією мовою". Технічні правила передачі інформації в мережі називають протоколами.

Протокол – це спосіб взаємодії, обміну даними між комп'ютерами під час роботи у мережі.

Сьогодні найбільш широко використовуються протоколи групи TCP/IP (англ. Transmission Control Protocol/Internet Protocol), і вся наукова інформація в Інтернеті передається за допомогою протоколів саме цієї групи.

Окрім того, є ще низка ще не стандартизованих, але доволі популярних протоколів – це протоколи електронної пошти (POP3, SMTP, IMAP), децентралізованого обміну файлами і текстовими повідомленнями; на деяких з них побудовано цілі файлообмінні мережі (OSCAR, CDDB, eDonkey, BitTorrent, Gnutella, Skype) та ін.

Розглянемо найбільш популярні протоколи, за якими здебільшого можна знайти наукову інформацію.

9.2.1. HTTP

Найпоширенішим протоколом є HTTP (Hyper Text Transfer Protocol) – протокол передачі гіпертекстових документів. Основним призначенням протоколу HTTP є передача веб-сторінок (текстових файлів з розміткою HTML), хоча з його допомогою успішно передають й інші файли, які пов'язані з веб-сторінками (зображення, відео, скрипти та ін.), які не пов'язані з ними (у цьому HTTP конкурує з складнішим FTP).

HTTP припускає, що клієнтська програма – веб-браузер – здатна відображати гіпертекстові веб-сторінки і файли інших типів у зручній для користувача формі. Для правильного відображення HTTP дає змогу клієнтові дізнатися мову і кодування веб-сторінки і/або запитати версію сторінки в потрібних мові/кодуванні, використовуючи позначення зі стандарту MIME.

HTTP – протокол прикладного рівня, подібними до нього є FTP і SMTP. Обмін повідомленнями відбувається за звичайною схемою "запит-відповідь". Для ідентифікації ресурсів HTTP використовує глобальні URL. На відміну від багатьох інших протоколів, HTTP не зберігає свого стану. Це означає, що сервер "не пам'ятає", хто звертався до нього, і про що йшлося мілісекунду тому. Саме тому компоненти, що використовують

HTTP, повинні самостійно зберігати інформацію про стан, пов'язаний до запитами і відповідями. Браузер, що посилає запити, може відстежувати затримки відповідей. Сервер може зберігати IP-адреси і заголовки запитів останніх клієнтів.

9.2.2. FTP

FTP (англ. File Transfer Protocol – протокол передачі файлів) – дає абонентові можливість обмінюватися бінарними та текстовими файлами з будь-яким комп'ютером мережі, що підтримує протокол FTP. Установивши зв'язок із віддаленим комп'ютером, користувач може скопіювати з нього файл на свій комп'ютер або скопіювати файл зі свого комп'ютера на віддалений.

Розглядаючи FTP як сервіса Інтернет, мають на увазі не просто протокол, а саме сервіс – доступ до файлів, які розміщені у файлових архівах.

FTP – стандартна програма, яка працює за протоколом TCP, котрий завжди встановлюють разом з операційною системою. Її початкове призначення – передача файлів між різними комп'ютерами, які працюють у мережах TCP/IP: на одному з комп'ютерів працює програма-сервер, на іншому – програма-клієнт, запущена користувачем, яка з'єднується із сервером і передає або отримує файли через FTP-сервіс. Звичайно для роботи через цей протокол користувачу потрібно бути зареєстрованим на FTP-сервері та використовувати логін і пароль для доступу.

Ця особливість і була причиною того, що програми FTP стали частиною окремого сервісу Інтернет. Справа в тому, що доволі часто сервер FTP налаштовують так, що з'єднатися з ним можна не тільки під своїм ім'ям, а й під умовним іменем anonymous (анонім). Тоді для користувача стає доступною не вся файлова система комп'ютера, а лише деякий набір файлів на сервері, які належать до вмісту сервера anonymous FTP – публічного файлового архіву. Сьогодні публічні файлові архіви організовано переважно як сервери anonymous FTP. Перелік інформації, яка міститься на таких серверах, охоплює всі аспекти життя: від звичайних текстів до мультимедіа.

Попри свою поширеність, у FTP є багато недоліків: програми-клієнти FTP не завжди зручні й прості в користуванні; користувач не завжди може зрозуміти, який файл перед ним:

той, що потрібен, чи ні. Окрім того, немає простого й універсального засобу для пошуку на серверах anonymous FTP (хоча для цього й існує спеціальний сервіс archie, але це незалежна програма, вона не універсальна і не завжди її можна ефективно застосовувати). Програми FTP доволі застарілі і деякі їхні особливості, які були потрібні при їх створенні, не зовсім зрозумілі і не потрібні сьогодні. Наприклад, для передачі файлів існує два режими – двійковий та текстовий, і, якщо користувач неправильно обрав режим передачі, то файл, який потрібно передати, може пошкодитися. Опис файлів на сервері існує у форматі операційної системи сервера, а список файлів операційної системи UNIX не завжди з розумінням сприймають користувачі Windows. Сервери FTP нецентралізовані, – звідси випливають ще деякі проблеми. Однак, незважаючи на все це, сервери anonymous FTP сьогодні – це стандартний спосіб організації публічних файлових архівів в Інтернеті.

9.2.3. Gopher

Gopher (англ. Gopher – суслик) – протокол передачі даних, який розробили у 1991 р. науковці університету Міннесоти (США). З 1993 р. почав втрачати свою популярність через швидкий розвиток HTTP. Він використовували для одержання інформації (переважно текстової) з ієрархічної файлової структури. Gopher був провісником WWW, що дав змогу за допомогою меню пересуватися від однієї сторінки до іншої, поступово звужуючи коло інформації, що відображається. Програми-клієнти Gopher мали текстовий інтерфейс.

Сервери Gopher широко використовували для поширення наукової інформації. Зараз більшість їх ресурсів стали доступними за допомогою протоколу HTTP.

Попри відносну застарілість у світі ще існує декілька серверів Gopher та певне коло їх користувачів.

Звертатися до ресурсів Gopher можна за допомогою браузера Firefox або через gopher-проксі (http://gopher.floodgap.com/gopher/gw).

9.3. Адресація в Інтернеті

Адреси потрібні для ідентифікації об’єктів, які можуть цікавити користувача в мережі. Найчастіше такими об’єктами є вузли мережі (сайти), поштові скриньки, файли, веб-сторінки.

Для кожного з них є власний формат адреси. Однак, оскільки об'єкти зосереджено у вузлах мережі, в їхніх ідентифікаторах обов'язково наявна адреса вузла.

Для ідентифікації вузлів слугує IP-адреса.

IP-адреса – це чотирибайтне число, що вказує на адресу комп'ютера в мережі.

Записують IP-адресу у вигляді чотирьох десяткових чисел (від 0 до 255), розділених крапками.

Приклад 13. *IP-адреса службового комп'ютера Супруновича С. В.:*

194.44.187.135 (у байтах у шістнадцятковому представленні: C2.2C.BB.87)

Перших два байти визначають адресу підмережі, а два інших – адресу вузла в ній. За допомогою IP-адреси можна ідентифікувати понад 4 млрд вузлів. На практиці ж через особливості адресації кількість можливих адрес становить понад 2 млрд. Фахівці передбачають, що у зв'язку з розширенням Інтернету всі можливі IP-адреси будуть використані до кінця 2011 р. – початку 2012 р. Тому зараз активно впроваджується новий стандарт IP-адреси – IPv6, в якому довжина адреси становить 16 байт ($\sim 3{,}4 \cdot 10^{38}$ можливих адрес).

Для користувача працювати із числовим зображенням IP-адреси незручно, тому було розроблено систему доменних імен DNS (англ. Domain Name System), котра могла перетворювати ім'я, записане звичайною англійською мовою, в IP-адресу.

Розподілену базу даних DNS підтримують за допомогою ієрархії DNS-серверів, що взаємодіють за однойменним протоколом. Основою DNS є подання доменного імені у вигляді ієрархічної структури. Кожен сервер, що відповідає за ім'я, може делегувати відповідальність за подальшу частину домену іншому серверові (з адміністративного погляду – іншій організації або людині), що дає змогу покласти відповідальність за актуальність інформації на сервери різних організацій, які відповідають тільки за підпорядковану їм частину доменного

імені.

Домен (від фр. domaine – область або одиниця структури) – певна зона в системі Інтернету, виділена власникові домену.

Доменне ім'я – символьне ім'я домену.

Це ім'я має бути унікальним у межах одного домену.

Контролюють використання доменних імен компанії-реєстратори. Вони мають повноваження створювати (реєструвати) нові доменні імена й продовжувати термін дії вже наявних доменних імен у доменах, для яких установлено обов'язкову реєстрацію. Такими доменами є: домен нульового рівня (кореневий домен); усі домени верхнього (першого) рівня; деякі домени другого рівня (наприклад, com.ua або co.uk). У всіх інших доменах для створення піддоменів особливі повноваження не потрібні. Як правило, реєстрація в доменах є платною – річна вартість домену може становити до 10 тис. доларів США, хоча зазвичай вона досить низька, а деколи буває й безкоштовна.

Усіма питаннями регулювання діяльності компаній-реєстраторів опікується Міжнародна некомерційна організація ICANN (англ. Internet Corporation for Assigned Names and Numbers – Інтернет-корпорація для призначення імен та номерів).

Реєстрація доменного імені не означає наявності сайту в Інтернеті, власник імені має знайти сервер для розміщення сайту. Є дуже багато імен, зареєстрованих зі спекулятивною метою – для подальшого перепродажу.

9.3.1. Структура доменних імен

Структура доменних імен повторює ієрархічну систему DNS та відповідає системі делегування повноважень компаній-реєстраторів (рис. 14).

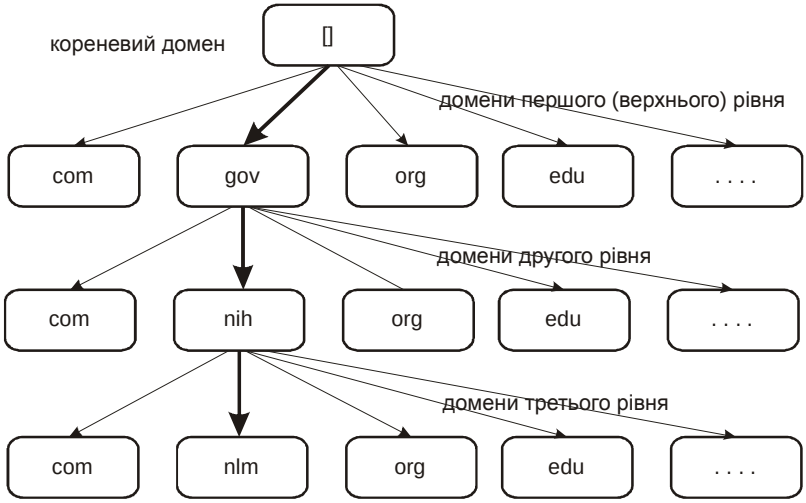


Рис. 14. *Ієрархічна структура доменних імен на прикладі http://pubchem.ncbi.nlm.nih.gov/*

9.3.2. WWW – Word Wide Web

Префікс www для написання доменного імені сьогодні вже не є технічною потребою. Раніше, коли протокол HTTP не був такий поширений, було вирішено для кожного доменного імені виділити субдомен www, який підкреслював, що ви перебуваєте на веб-сервері мережевого ресурсу, вказаного в імені домену. Незабаром ця технологія стала не тільки дуже популярною, а й заполонила мережу, тому сенсу використовувати цей субдомен вже не було.

Оскільки людина – дуже консервативна істота і важко сприймає всі зміни (адже тільки звикли, що www – це Інтернет), деякі веб-майстри створюють субдомен www, який так само за замовчуванням стає синонімом до основного доменного імені.

Сьогодні вже не тільки немає потреби у використанні префікса, а й навіть можна говорити про те, що він створює певну плутанину через те, що одна частина сайтів сприймає адреси із www, і без нього, друга частина без www не працює, а третя – не сприймає адреси з www.

Звідси випливає висновок, що коли є проблеми з доступом до сайту за адресою, варто спробувати обидва варіанти

написання адреси – з www та без, оскільки передбачити, які налаштування встановив веб-майстер сайту, практично не можливо.

9.3.3. Система адрес у мережі Internet

Ім'я та IP-адреса не тотожні – одна IP-адреса може мати кілька імен, що дозволяє підтримувати на одному комп'ютері декілька веб-сайтів (це називається віртуальний хостинг); водночас одному імені може відповідати декілька IP-адрес – це значно розширює можливості масштабування Інтернету.

9.3.4. Структура веб-адрес

Для ідентифікації ресурсів мережі (файлів, веб-сторінок) використовують адресу URL (англ. Uniform Resource Locator – уніфікований покажчик ресурсу), яка має таку структуру:

[схема://][логін:пароль@]хост[:порт]/[шлях] ,

де:

схема – схема звернення до ресурсу, у більшості випадків – це мережевий протокол, ftp://, http:// чи інший. У більшості браузерів протокол http:// може бути пропущений.

Логін та пароль для доступу до ресурсу є обов'язковими для електронної пошти й FTP. Більшість сучасних програм самостійно вставляють ці дані, використовуючи відповідні налаштування в URL при зверненні до інших серверів. Багато наукових веб-ресурсів також потребують цих даних, котрі можуть бути вставлені в URL. Якщо для входу в сервіс потрібен логін та пароль, а вони не задані, то, як правило, виводиться вікно для введення відповідних значень. Для ресурсів, що видають неоднакову інформацію різним користувачам, зазвичай використовують інші схеми ідентифікації, тож явне внесення логіна і пароля в URL не проходить.

Хост – IP-адреса сайту у формі чотирьох десяткових чисел, розділених крапками; числа – натуральні в інтервалі від 0 до 255, або доменне ім'я.

Порт – порт для підключення. Для HTTP – це, як правило, "80", і його пропускають. Якщо з'єднання відбувається за допомогою іншого порту, то його вказують.

Шлях – уточнювальна інформація про місцезнаходження ресурсу; залежить від протоколу та способу організації інформаційного обслуговування сайту. Зазвичай відображає структуру директорій на сервері, хоча дедалі частіше вказує не логічну структуру підпорядкування інформації.

Веб-адреси можуть бути двох типів – статичні та динамічні.

Статична веб-адреса – вихідний текст веб-сторінки, показаної у браузері, ідентичний вихідному текстові сторінки на сервері.

Приклад 14. *Статична веб-адреса:*

http://www.chem.isu.ru/eos/index.html

Статичні веб-сторінки, як правило, мають розширення HTM або HTML.

Динамічна веб-адреса – вихідний текст веб-сторінки, показаної у браузері, не відповідає вихідному текстові сторінки на сервері.

Приклад 15. *Динамічна веб-адреса:*

http://www.ximicat.com/info.php?id=798

Динамічні сторінки зазвичай формує певна програма, встановлена на сервері, й здебільшого сторінка, котру бачить користувач на екрані свого комп'ютера, існує в одиничному екземплярі – тільки на комп'ютері користувача. Хоча й тут можливі варіанти. Здебільшого на динамічність веб-сторінки вказує її розширення, хоча технології Інтернету дають змогу зробити динамічною сторінку з будь-яким розширенням. Динамічні сторінки найчастіше мають розширення PHP, ASP, CGI, EXE, PL та ін. З огляду цитування веб-адрес можливі два випадки: перший – коли в адресному рядку URL передаються певні параметри, тоді будь-хто з будь-якого підключеного до Інтернету комп'ютера, якщо введе таку адресу в браузер, одержить ту саму сторінку.

Приклад 16. *Стаціонарна динамічна веб-адреса:*

http://www.chemport.ru/pertable/elinfo.php?el=5

Другий випадок – коли використовують куки та сесії. URL тоді дійсний тільки обмежений період часу і тільки на

комп'ютері того користувача, який відкрив ту сторінку, і через певний період або після закриття вікна браузера стає недійсним. Ознака такої адреси – випадкова послідовність цифр і букв.

Приклад 17. *Нестаціонарна динамічна веб-адреса:*

http://upload.com.ua/dget/900436492/?hash=5973b7b144e1b031476a927892ba4b64&shsid=41228262

На такого виду адреси ні в якому разі посилатися не можна. Якщо ж треба послатися на такий ресурс, то використовують аналітичний бібліографічний опис.

URL інших ресурсів здебільшого статичні. Найчастіше виникає потреба користуватися електронною поштою та FTP.

Приклад 18. *Адреса електронної пошти:*

kormosh@univer.lutsk.ua

Приклад 19. *FTP-адреса:*

ftp.tex.ac.uk

9.4. Питання для самоперевірки

1. Що таке Інтернет?
2. Яка роль системи DNS?
3. Що таке мережевий протокол?
4. Які найпоширеніші протоколи Інтернету?
5. Для чого використовують протокол HTTP?
6. Призначення протоколу FTP.
7. Що таке IP-адреса?
8. Доменне ім'я – це...
9. Структура доменного імені.
10. Яка загальна структура веб-адрес?
11. Які види адрес застосовують в Інтернеті?
12. Які види адрес не можна використовувати як посилання і чому?

9.5. Структура Інтернету

Сьогодні до Інтернету входить низка найрізноманітніших сервісів, або служб, кожна з яких має власні особливості. Зараз найпопулярнішими службами Інтернету є: Веб, Веб-форуми, Блоги, Вікі-проекти (включно з Вікіпедією), інтернет-магазини,

інтернет-аукціони, інтернет-новини, електронна пошта та списки розсилки, групи новин (в основному, Usenet), файлообмінні мережі, електронні платіжні системи, інтернет-радіо, інтернет-телебачення, IP-телефонія, системи обміну повідомленнями, FTP-сервери, IRC, RSS.

З огляду на доступ до наукової інформації найбільш корисними є веб. Інші служби можна використовувати хіба що для підвищення ефективності наукової роботи завдяки прискоренню обміну різноманітною інформацією.

Практично всі послуги мережі Інтернет побудовані за принципом "сервер–клієнт".

Сервер (у мережі Internet) – це комп'ютер або програма, здатні надавати клієнтам (у міру надходження від них запиту) деякі послуги мереж.

Клієнт – прикладна програма, завантажена в комп'ютер користувача, яка забезпечує передачу запитів до сервера й одержання відповідей від нього.

Різні сервіси мають неоднакові прикладні протоколи. У міру розвитку мережі з'являються нові протоколи (сервіси), змінюючи її вигляд і стрімко розширюючи коло користувачів. Так, щоб скористатися однією зі служб мережі, потрібно встановити на комп'ютері клієнтську програму, здатну працювати за протоколом цієї служби. Деякі клієнтські програми входять до складу операційних систем, або до складу програм-браузерів, наприклад, MS Internet Explorer, MS Outlook Express. Розглянемо деякі сервіси, які забезпечує Інтернет.

9.6. Класифікація сервісів Інтернету

9.6.1. FTP

Сервіс FTP (File Transfer Protocol) – це протокол передачі файлів, один із перших сервісів Інтернету, який дає можливість абоненту обмінюватися двійковими і текстовими файлами з будь-яким комп'ютером мережі. Встановивши зв'язок із віддаленим комп'ютером, користувач може скопіювати файл із нього на свій комп'ютер або зі свого на віддалений. Для вузлів

FTP характерна наявність процедури входу (login). Як “гостьові” ім’я і пароль часто використовують імена anonymous, ftp, а також адресу електронної пошти, при цьому користувач має доступ до безкоштовно поширеної інформації. Для зручності роботи з цим сервісом розроблено низку прикладних програм, що забезпечують зручний Windows-подібний інтерфейс для FTP-сервісу. Цей сервіс можна використовувати для комерційного поширення програмних продуктів, баз даних, моделей, рекламних презентацій, великих за обсягом документів (книг) тощо.

9.6.2. E-mail

Електронна пошта (E-mail) – це офлайн-служба, яка забезпечує обмін поштовими повідомленнями з будь-яким абонентом мережі Інтернет. Можна відправити і текстові, й бінарні файли. Вона є одним із перших і, мабуть, найпоширеніших сервісів. Електронна пошта – це найдешевший і доступний Інтернет-сервіс в Україні. Можна вказати на такі переваги електронної пошти в організації ділової діяльності:

– поштові сервери одержують повідомлення від клієнтів і пересилають їх по ланцюжку до поштових серверів адресатів, де ці повідомлення накопичуються. При встановленні сполучення між адресатом і його поштовим сервером за командою відбувається передача повідомлень, що надійшли на комп’ютер адресата. Серед клієнтських поштових програм можна назвати The Bat, Microsoft Outlook та інші.

9.6.3. Списки розсилки

Сервіс Mail Lists (списки розсилки) використовує поштові протоколи для розсилки повідомлень. Підписавшись на списки розсилки, можна електронною поштою регулярно одержувати повідомлення про певні теми (науково-технічні й економічні огляди, презентації нових програмних та апаратних засобів тощо).

Зараз підписку мають майже всі великі сайти в Інтернеті. Щоб бути в курсі досягнень хімічної науки, варто підписатися на новини декількох ключових хімічних порталів.

Зворотна сторона списків розсилки – небажана реклама, або СПАМ (SPAM – англ. SPiced hAM, гостра шинка). Назва

виникла 1936 р., коли американська компанія Hormel Foods проводила рекламну компанію м'ясних консервів. Відтоді спамом почали називати всі небажані рекламні розсилки. Коли електронна адреса потрапляє в бази спамерів, у поштову скриньку починає надходити все більше й більше непотрібних рекламних листів. Тому треба бути обережним перед тим, як вказувати адресу поштової скриньки на сумнівних Інтернет-сайтах. Краще мати декілька поштових адрес різного призначення, тим більше, що зараз створити безкоштовну поштову скриньку можна дуже швидко.

9.6.4. Usenet

Сервіс Usenet (групи новин або телеконференції) забезпечує обмін інформацією (повідомленнями, статтями) між усіма, хто користується ним. Це щось на зразок електронної дошки оголошень, на яку будь-хто може розмістити своє повідомлення, і воно стає доступним для всіх інших. Новини поділяють за темами на групи, що певною мірою сприяє їхньому упорядкуванню. На окремі з них можна оформити підписку, й періодично, як і електронною поштою, одержувати всі повідомлення, що з'являються на тему обраної групи. Існують групи новин, що розповсюджують повідомлення про наукову інформацію. Для реалізації цього сервісу існують клієнтські програми, наприклад, Microsoft Internet News. Зараз цей сервіс малопоширений.

9.6.5. WWW

Сервіс WWW (World Wide Web – всесвітня павутина) – це єдиний інформаційний простір, який складається із сотень мільйонів взаємопов'язаних гіпертекстових електронних документів, що зберігаються на Web-серверах. Окремі документи всесвітньої павутини називаються Web-сторінками, групи тематично об'єднаних Web-сторінок утворюють Web-сайт, або просто сайт. Web-сторінка – це текстовий файл, що містить опис зображення мультимедійного документа мовою гіпертекстової розмітки – HTML (Hyper-Text Markup Language). Сторінка може містити не тільки форматований текст, а й графічні, звукові та відеооб'єкти.

Найважливішою рисою Web-сторінок є гіпертекстові

посилання. З будь-яким фрагментом тексту або, наприклад, із малюнком, можна пов'язати інший Web-документ, тобто встановити гіперпосилання. Тоді, натиснувши лівою клавішею миші на текст або рисунок, що є гіперпосиланням, відправляється запит на доставку нового документа. Цей документ, своєю чергою, також може мати гіперпосилання на інші документи. Отож, сукупність величезної кількості гіпертекстових електронних документів, які зберігаються у серверах WWW, утворює своєрідний гіперпростір документів, між якими можливе переміщення.

Для передачі інформації у WWW використовують протокол HTTP. Перегляд Web-сторінок і переміщення за посиланнями здійснюють за допомогою програм браузерів (від слова "to browse" – переглядати). Найпопулярнішими Web-браузерами в Україні є Microsoft Internet Explorer та Mozilla Firefox.

9.6.6. Telnet

Telnet – найстаріший сервіс Інтернет. За його допомогою можна "потрапити" на віддалений комп'ютер мережі, запустити на ньому програму або подивитися зміст певного файлу. Він дає можливість абонентові працювати на будь-якому комп'ютері мережі Інтернет, як на власному. Часто використовують режим роботи "доступ до віддаленого сервера бази даних". Засобами Telnet можна досліджувати величезні бази даних і навіть бібліотеки в усьому світі, щоб перевірити наявність потрібного ресурсу, можна потрапити у простори, недоступні у Web.

У минулому цей сервіс також широко використовували для складних математичних розрахунків на віддалених суперкомп'ютерах. У наші дні, через швидке підвищення обчислювальної потужності ПК, потреба у подібній послузі скоротилася, але служби Telnet у мережі Інтернет існують і сьогодні. Часто протоколи Telnet застосовують для дистанційного керування технічними об'єктами, наприклад, телескопами, відеокамерами, промисловими роботами.

Варто зазначити, що Telnet, як правило, працює у режимі командного рядка, тобто не має графічного інтерфейсу, тому ним може послуговуватися тільки досвідчений користувач. Без суттєвої попередньої підготовки доступ за його допомогою до

наукових ресурсів не буде можливим.

У комп'ютерній мережі ВНУ ім. Лесі Українки доступ до зовнішніх серверів за допомогою Telnet закритий!

9.6.7. Електронні дошки оголошень

Електронна дошка оголошень (англ. Bulletin Board System, BBS) – як правило, так називають невеликі системи, призначені для місцевих користувачів. Сучасні BBS трапляються і використовуються рідко. За кордоном BBS використовують, зокрема, для розповсюдження інформації про організації, місцеві події, літературу (книжкові огляди), що випускають видавництва, та ін.

9.6.8. Веб-конференції

Електронна конференція, що проводиться в мережах Інтернету, подібна до автоматично підтримуваних списків розсилки електронних листів за певними тематиками. У межах теми конференції абонент пише лист, надсилає його на певну адресу, і лист автоматично розсилається всім, хто підписався на цю конференцію. Основна відмінність від списків розсилки полягає в тому, що повідомлення розсилаються не конкретному користувачеві, а зберігаються на багатьох серверах, спеціально організованих для підтримки відповідних конференцій. Повідомлення через деякий час видаляються. Такі конференції існують доти, поки є люди, які в них пишуть.

9.6.9. Веб-форуми

Форуми утворюють в Інтернеті співтовариства людей (групи користувачів глобальної мережі) для обміну думками, консультування, отримання новин за інтересами. Вони більше схожі на записи розмов, які деколи зберігаються в мережі досить довго. Для участі в них потрібно зареєструватися, надавши певну інформацію про себе. Як правило, на форумах зрідка можна знайти наукову інформацію. Проте, оскільки форуми – це один із видів спілкування в мережі, за їх допомогою можна одержати оцінку певної інформації, рецензії, цінні поради, що корисні для наукової діяльності, зрештою, одержати відповідь на питання або посилання на давно шукане джерело інформації.

Форуми – це вид спілкування в мережі, а для успішного їх

використання варто дотримуватися нескладних звичних правил етики, як і в повсякденному житті. Крім того, учасники форуму можуть стати добровільними помічниками при розв'язанні складних і не наукових проблем. Єдине, що для цього треба – заслужити хорошу репутацію на відповідному форумі, яку досягають завдяки постійній участі у спілкуванні та допомозі, наданій іншим учасникам форуму.

9.6.10. IRC

Сервіс IRC (IRC, англ. Internet Real Chat – реальна розмова в Інтернеті) – служби обміну текстовими повідомленнями, що дають змогу організувати "розмову" двох або більше користувачів у режимі реального часу. Користувачі інтерактивно спілкуються один з одним за допомогою клавіатурного введення текстів у спеціальному вікні програми. Це своєрідна розмова між людьми у письмовій формі в мережі Інтернет.

На відміну від форумів, інформація на чатах довго не зберігається, й через невеликий проміжок часу стає не доступною. Щодо наукової інформації, якщо це спеціалізований науковий чат, то тут зазвичай можна досить швидко з'ясувати окремі питання.

9.6.11. ICQ

Різновидом IRC є ICQ (ICQ, від англ. I Seek You – я шукаю тебе) сервіс, який використовують для інтерактивного спілкування двох користувачів Інтернету, автоматично викликає і показує, хто зі знайомих перебуває на зв'язку. Кожен учасник має власний ідентифікаційний номер.

9.6.12. Відеоконференції

Різновидом електронної конференції є телеконференція, або відеоконференція – засіб і спосіб візуального інтерактивного спілкування у комп'ютерних мережах територіально віддалених людей. Як правило, для їх проведення потрібно підключити до комп'ютера мікрофон та веб-камеру, щоб передавати на відстань голос і зображення. Виділяють три варіанти проведення відеоконференцій: студійний, груповий і персональний.

У відеоконференціях користувачам забезпечується виділений доступ до інформації групового використання для суміс-

ного проведення електронних тематичних конференцій, при цьому кожен учасник бачить і чує своїх співрозмовників.

Аудіоконференції (голосові конференції) забезпечують користувачам в Інтернеті виклик, з'єднання і розмову аналогічно до звичайної телефонної розмови.

9.6.13. Інтернет-телефонія

Інтернет-телефонія (англ. Internet Telephony – IT і IP-телефония) надає можливість розмовляти по телефону, використовуючи технології Інтернету.

За допомогою комп'ютерної техніки і телекомунікацій вона в реальному часі забезпечує недорогий голосовий зв'язок. IP-телефонія дає змогу використовувати будь-яку IP-мережу для ведення міжміських та міжнародних телефонних розмов і передачі факсів у режимі реального часу.

9.6.14. Інтернет-телебачення

Інтернет-телебачення передбачає трансляцію через Інтернет відеоматеріалу телевізійної якості зображення і звуку. Відмінність від відеоконференції полягає в односторонній передачі потокового зображення і звуку.

9.6.15. Пірингові мережі

Пірингові (однорангові, децентралізовані, від англ. peer-to-peer – рівний до рівного) мережі – це комп'ютерні мережі, засновані на рівних правах учасників. У таких мережах немає виділених серверів, а кожен вузол (peer) є і клієнтом, і сервером. На відміну від архітектури, клієнт-сервер – така організація зберігає працездатність мережі при будь-якій кількості і будь-якому поєднанні доступних вузлів.

Першу пірингову мережу запустив 1984 р. Парбауелл Йохнухуітсман (Parbawell Yohnuhuitsman), розробляючи архітектуру Advanced Peer to Peer Networking фірми "IBM".

Найпоширенішими піринговими мережами є ED2K (eDonkey2000), BitTorrent, Gnutella, Skype та інші. Ці мережі можуть бути надзвичайно корисними при одержанні недоступних наукових матеріалів. Недоліком є те, що наповнення цих мереж часто не індексують пошукові системи, та ці прогресивні технології постійно конфліктують з організаціями, що слідкують за забезпеченням авторських прав. Це ще один із

яскравих прикладів, коли застарілі суспільні відносини гальмують розвиток сучасних технологій.

У науці технологія пірінгових мереж тією чи іншою мірою застосовується також для розподілених обчислень.

9.6.16. Проекти розподілених обчислень

Розподілені обчислення дають змогу за порівняно короткі терміни виконувати досить значний обсяг розрахунків, який навіть на суперкомп'ютерах вимагав би, залежно від складності завдання, багатьох років або навіть століть роботи. Такої продуктивності досягають завдяки тому, що певне глобальне завдання розбивають на велику кількість блоків, які одночасно виконують сотні чи тисячі комп'ютерів. Як правило, такі розрахунки проводять на персональних комп'ютерах добровольців, котрі бажають таким способом зробити внесок у розвиток науки.

Із відомих проектів розподілених розрахунків у хімії можна відзначити такі:

– Spinhenge@home

Spinhenge@home (http://spin.fh-bielefeld.de) досліджує наномагнітні молекули. Передбачається, що ці дослідження сприятимуть прогресу в розробці ліків для хіміотерапії та мініатюрних елементів пам'яті.

– Hydrogen@Home

Hydrogen@Home (http://hydrogenathome.org) вивчає методи виробництва та зберігання водню задля розробки екологічно чистих технологій.

– Predictor@home

Predictor@home (http://predictor.scripps.edu) використовує розподілені обчислення для передбачення структури білка за даними про послідовність його атомів. Робота проекту спрямована на перевірку нових алгоритмів і методів передбачення структур білків. Мета – удосконалити відомі алгоритми передбачення, а також знайти відповідь на найважливіші питання, пов'язані з хворобами, які викликані порушеннями у білкових структурах.

– Docking@Home

Docking@Home (http://docking.cis.udel.edu) – спільний некомерційний проект розподілених обчислень Університету Делавера, Науково-дослідного інституту Скріппса та Університету Каліфорнії в Берклі, які підтримує Національний науковий фонд США. У проекті попередньо відбирають речовини, які можуть стати основою нових або поліпшення вже відомих лікарських препаратів. Для цього комп'ютери учасників прораховують хімічні властивості комплексів ліганд-білок (мала молекула-рецептор) – цей процес ще називають молекулярним докінгом.

9.7. Формати файлів з Інтернету

Інформація, що подається нам за допомогою браузера, має формат HTML (англ. HyperText Markup Language – мова розмітки гіпертексту) – стандартна мова розмітки документів у Всесвітній мережі. Більшість веб-сторінок створюються за допомогою мови HTML. Документ HTML оброблює браузер та відтворює на екрані у звичному для людини вигляді. Проте для представлення наукової інформації здебільшого використовують інші формати. Їх можна умовно розділити на дві групи: перша – це представлення наукових документів; друга група – це представлення винятково хімічної інформації у цифровій формі.

9.7.1. Електронні документи

Електронні документи подають здебільшого у форматах PDF, DjVu. Головним завданням при розробці цих форматів було найбільш точно передати вигляд реального, друкованого документа. З часом сфера застосування цих форматів значно розширилася, тепер часто електронні документи не мають паперового першоджерела, але за наявності потрібної техніки їх можна легко перетворити в паперові з мінімальними змінами у змісті та оформленні.

9.7.2. Формат PDF

PDF (англ. Portable Document Format – формат переносних документів) – багатоплатформовий формат електронних документів, які створила компанія Adobe Systems з використанням багатьох можливостей мови Postscript. Передусім, він призначений для представлення в електронному вигляді поліграфічної продукції, –

значна кількість сучасного професійного друкарського устаткування може обробляти PDF безпосередньо. Для перегляду можна використовувати офіційну безкоштовну програму Adobe Reader, а також програми сторонніх розробників, такі як Foxit Reader. Кожен PDF-файл може містити повну інформацію про документ, наприклад, тексти та зображення, крім того у файлі можуть міститися елементи, неможливі для друкованого документа, такі як векторні зображення, відео, інтерактивні форми та ін. У грудні 2007 р., формат PDF затверджено як стандарт ISO 32000.

Статті наукових журналів, доступних в електронному вигляді, як правило, мають цей формат.

9.7.3. Формат DjVu

DjVu (фр. déjà vu – колись уже бачене) – технологія стискання зображення із втратами. Розроблене компанією AT&T спеціально для зберігання відсканованих документів – книг, журналів, рукописів та ін., де наявна велика кількість формул, схем, рисунків та рукописних символів, котрі ускладнюють повноцінне розпізнавання такого документа. Також це дуже ефективне рішення, якщо потрібно виразити всі особливості його оформлення. Наприклад, в історичних документах важливі не тільки зміст, а й колір, фактура паперу, його дефекти: тріщини, сліди від згинів, плями, залишені сліди предметів тощо.

DjVu став основою для декількох бібліотек наукових книг. Величезна кількість літератури у цьому форматі доступна у файлообмінних мережах. Формат оптимізований для передачі у мережі так, що сторінку можна переглядати ще до завершення стягування інформації. DjVu-файл може містити текстовий (OCR) шар, що дає змогу здійснювати повнотекстовий пошук по файлу. Крім того, DjVu-файл може містити вбудований інтерактивний зміст й активні ділянки – посилання, що дають можливість реалізувати зручну навігацію за змістом книги.

Як вже було сказано, у форматі DjVu зберігають, в основному, відскановані книги. Файли у ньому значно компактніші, ніж у PDF. Пошук потрібних книг у цьому форматі дещо ускладнений через те, що вміст піринговых мереж не завжди індексується пошуковими системами, або книги часто недостатньо описані для швидкої та чіткої ідентифікації. Електронні DjVu-бібліотеки зазнають переслідування з боку організацій, що слідкують або володіють

авторськими правами на книги.

Для перегляду документів у форматі DjVu потрібні спеціальні програми, або плагіни до браузерів.

9.7.4. Формати файлів, що містять винятково інформацію з хімії

Сюди належать спеціалізовані формати структурної, спектральної або іншої інформації.

9.7.5. Формат MOL

Загальноприйнятий формат представлення і плоскої, й об'ємної структури молекул. Розширення файлів – MOL.

9.7.6. Формат SVG

Векторний формат, що набуває все більшої популярності. Його застосовують, в основному, для демонстрування спектрів та іншої графічної інформації. Поширений у Вікіпедії. Розширення файлів – SVG.

9.7.7. Формат JCAMP-DX

JCAMP-DX (англ. the Joint Committee on Atomic and Molecular Physical data – Data Exchange format, – Міжвідомча комісія з атомних і молекулярних фізичних даних – формат обміну даними) – стандартна форма файла для обміну ІЧ, ЯМР і мас-спектрів та пов'язаної з ними інформації з хімії та фізики між системами даних спектрометрів різних виробників. Це текстовий формат. Розширення файлів – DX або JDX.

9.7.8. Формат PDB

PDB (англ. Protein Data Bank – банк даних протеїну) – формат запису структури протеїнів, але його можна використовувати й для інших типів молекул. Оскільки у цьому форматі можна описати комплекси макромолекул, то ці файли можуть досягати значних розмірів, тому частіше їх подають у заархівованому вигляді. Типове їх розширення – PDB, хоча і трапляються застарілі розширення ENT або BRK.

9.8. Питання для самоперевірки

1. Що стало причиною створення Інтернету?
2. Які технічні досягнення уможливили існування Інтернету?
3. Що таке протокол Інтернету?

4. Назвіть основні протоколи Інтернету.
5. За яким протоколом передаються веб-сторінки?
6. За яким протоколом передаються файли?
7. Що таке IP-адреса?
8. Що таке доменне ім'я?
9. Назвіть основні сервіси Інтернету.
10. Які сервіси Інтернету забезпечують безпосередній доступ до наукової інформації?
11. Що таке пірингові мережі?
12. У чому суть проектів розподілених обчислень?
13. Які формати файлів використовують для розповсюдження документів?
14. Які формати файлів використовують для розповсюдження інформації з хімії?

10. ПОШУКОВІ МАШИНИ

У світі нараховують більше 45 тис. різних пошукових машин, які розрізняють за обсягом і спеціалізацією, поділяючи їх на загальні та спеціалізовані.

Без використання цих систем практично не можливо знайти щось корисне серед безлічі інформаційних ресурсів мережі. Але що вони собою являють, як влаштовані, чому результат пошуку в терабайтах інформації видається так швидко, як влаштовано ранжирування документів при видачі, що таке інформаційні масиви цих систем – цим питанням присвячений розділ.

Інформаційно-пошукові системи з'явилися достатньо давно. Теорію і практику конструювання таких систем започатковано з кінця 70 – поч. 80 рр. XX ст. Отже, не можна сказати, що з появою Інтернету і бурхливим входженням його в практику інформаційного забезпечення, з'явилося щось принципово нове, невідоме раніше. Якщо точніше, то інформаційно-пошукові системи в Інтернеті – це визнання того, що ні ієрархічна модель Gopher, ні гіпертекстова модель World Wide Web не вирішують проблему пошуку інформації у великих обсягах різнорідних документів. Хоча на сьогодні немає іншого способу швидкого пошуку даних, окрім пошуку за ключовими словами.

При використанні ієрархічної моделі Gopher доводиться досить довго шукати в каталогах потрібну інформацію. Ці каталоги хтось має підтримувати і при цьому їх тематичне

групування має збігатися з інформаційними потребами користувача. Зважаючи на анархічність Інтернету і величезну кількість усіляких інтересів у користувачів мережі, зрозуміло, що комусь може й не пощастити – в мережі не буде каталогу, який відображає конкретну предметну ділянку. Саме тому для великої кількості серверів Gopher розроблено інформаційно-пошукову програму Veronica (англ. Very Easy Rodent-oriented Net-wide Index of Computerized Archives – дуже легкий індекс комп'ютеризованих архівів, що охоплює мережу, орієнтований на гризунів).

Аналогічний розвиток подій ми бачимо й у World Wide Web. Ще у 1988 р. Франк Халаж (Frank G. Halasz) серед ключових проблем розробки гіпертекстових систем і їх використання назвав проблему організації пошуку інформації у великих гіпертекстових мережах як гловне завдання для наступного покоління систем цього типу. Природно, що система, яку запропонував Бернерс-Лі і яка набула значного поширення в Інтернеті, мала зіштовхнутися з тими самими проблемами, що і її локальні попередники. Реальне підтвердження цьому продемонстровано на другій конференції з World Wide Web восени 1994 р., на якій виголошено доповіді на тему розробки інформаційно-пошукових систем для Web, а система World Wide Web Worm, котру розробив Олівер Макбрайн з університету Колорадо, отримала приз як кращий навігаційний засіб. Також слід відзначити, що все-таки тривале існування й використання призначене не хорошим програмам окремих талановитих винахідників, а засобам, які є результатом довгострокового планування послідовного руху до поставленої мети наукових і виробничих колективів. Рано чи пізно етап досліджень закінчується і настає період експлуатації систем, а це вже зовсім інший вид діяльності. Саме така доля чекала два інших проекти, представлених на тій конференції: Lycos, що його підтримала компанія Microsoft, і Webcrawler, який став власністю America On-line.

Розробка нових інформаційних систем для Web не завершена, як на стадії написання комерційних систем, так і на етапі досліджень. Багато проблем, які ставить перед розробниками ІПС Інтернет, не вирішені й нині. Саме цією

обставиною і викликане виникнення проектів типу Altavista компанії Digital, головною метою якого є розробка програмних засобів інформаційного пошуку для Web і підбору архітектури для інформаційного сервера Web.

10.1.1. Архітектура сучасних інформаційно-пошукових систем World Wide Web

Перш ніж описати проблеми побудови інформаційно-пошукових систем Web і методи їх вирішення, розглянемо типову схему такої системи (рис. 15). У різних публікаціях, присвячених конкретним системам, подані схеми, які відрізняються одна від одної тільки застосуванням конкретних програмних рішень, але не принципом організації різних компонентів системи. Розглянемо цю схему на прикладі:

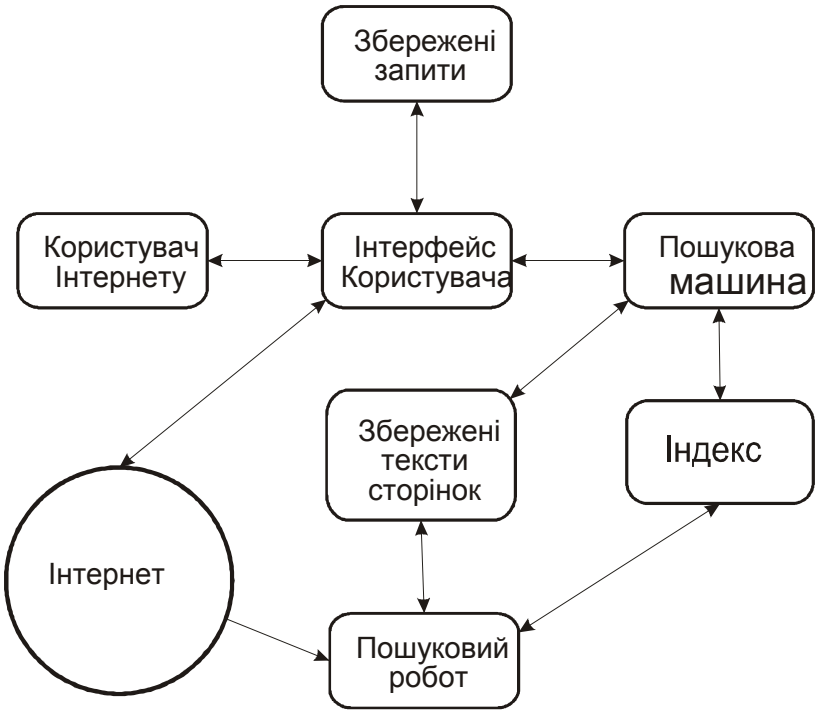


Рис. 15. *Структура інформаційно-пошукових систем*

На цій схемі позначені:

Користувач Інтернету – це людина (або програма), що переглядає конкретний інформаційний ресурс документів World Wide Web, Gopher, Wais, FTP-архівів, поштових списків розсилки і груп новин Usenet. Своєю чергою всі ці інформаційні ресурси є об'єктом пошуку інформаційно-пошукової системи.

Інтерфейс користувача – це комплекс програмного й апаратного забезпечення, що дає користувачеві Інтернету можливість взаємодіями з ресурсами та сервісами мережі.

Пошукова система слугує для трансляції запиту користувача, який готується інформаційно-пошуковою мовою (ІПМ), у формальний запит системи, пошуку посилань на інформаційні ресурси мережі й видачі результатів пошуку користувачеві.

Індекс – це основна кількість даних інформаційно-пошукової системи, він містить оброблену інформацію про ресурси мережі та забезпечує пошуковій системі швидкий пошук та видачу результатів.

Збережені запити користувача місьтяться в його особистій базі даних. Сюди ж належить інформація, одержана за допомогою інтерфейсу користувача.

Пошуковий робот призначений для сканування Інтернету й підтримки бази даних індексу в актуальному стані. Ця програма є основним джерелом відомостей про стан інформаційних ресурсів мережі.

Інтернет – ресурси мережі, що доступні за допомогою інтерфейсу користувача.

Збережені тексти сторінок пошукова система використовує для видачі розширених результатів пошуку, коли крім посилання на ресурс наводиться фрагмент тексту відповідного документа.

Розглянемо детальніше призначення і принцип побудови кожного з цих компонентів, й визначимо, в чому полягає відмінність цієї системи від традиційної інформаційно-пошукової системи локального типу.

10.1.2. Інформаційні ресурси та їх місце в інформаційно-пошуковій системі

Як видно зі схеми (рис. 15, с. 93), документальним масивом ІПС Internet є сукупність документів Інтернету. Усе це досить різнорідна інформація, яка представлена у вигляді різних, неузгоджених один з одним форматів даних. Тут є і текстова, графічна, й аудіоінформація та загалом усе, що містять вказані вище сховища. Постає питання: як інформаційно-пошукова система може зі всім цим працювати? У традиційних системах є поняття пошукового образу документа – ПОД – це певний елемент, що замінює собою документ і використовується під час пошуку замість реального документа. Пошуковий образ є результатом застосування певної моделі інформаційної сукупності документів до реальної їх кількості. Найбільш популярною моделлю є векторна, у якій до кожного документа додають список термінів, що найбільш точно відображають його зміст. Якщо точніше, то документові приписують вектор, обсяг якого дорівнює кількості термінів, якими можна скористатися при пошуку. У найпростішій векторній моделі елемент вектора дорівнює 1 або 0, залежно від наявності терміна в ПОД документа чи його відсутності. У складніших моделях терміни зважують, тобто елемент вектора дорівнює не 1 або 0, а деякому числові, яке відображає відповідність цього терміна документа. Саме остання модель найбільш популярна в інформаційно-пошукових системах Internet. Існують й інші моделі опису документів: імовірнісна модель інформаційних потоків і пошуку, а також модель пошуку в нечітких множинах. Аналіз переваг і недоліків застосування цих моделей при реалізації інформаційно-пошукових систем в Інтернет – це тема спеціального дослідження. Тут варто звернути увагу тільки на те, що поки у пошукових системах переважає саме лінійна модель. Дослідження із застосування інших моделей проводяться постійно, і час від часу в Інтернеті з'являються ресурси, у яких зроблено спробу реалізувати ту чи іншу модель. Отож, перше завдання, яке має вирішити інформаційно-пошукова система, – це приписування списку ключових слів документові або інформаційному ресурсові. Саме ця процедура і називається індексуванням. Часто, проте, індексуванням називають складання файла інвертованого списку, в якому кожному термінові індексування ставлять відповідний список документів, у яких він трапляється. Така процедура є лише

технічним аспектом створення пошукового апарату інформаційно-пошукової системи.

Проблема, пов'язана з індексуванням, полягає в тому, що приписування пошукового образу документові або інформаційному ресурсові спирається на уявлення про словник, із якого ці слова вибирають, як про фіксовану сукупність термінів. У традиційних системах практикують розбиття на системи з контрольованим словником і системи з вільним словником. Контрольований словник припускав ведення деякої лексичної бази даних, додавав терміни до якої адміністратор системи. Тому всі нові документи могли бути заіндексовані тільки тими термінами, які містила ця база даних. Вільний словник поповнювався автоматично у міру створення нових документів. Проте на момент актуалізації словник також фіксувався. Актуалізація припускала повне перезавантаження бази даних. При такому оновленні перезавантажувалися власне документи й оновлювався словник, а після цього проводилася переіндексація документів. Процедура актуалізації забирала досить багато часу і доступ до системи у момент її актуалізації закривався.

Тепер уявімо собі можливості такої процедури в анархічному Інтернеті, де ресурси з'являються і зникають щодня. При створенні програми Veronica для Gopherspace передбачали, що усі сервери мають бути зареєстровані і так вели облік наявності або відсутності ресурсу. Veronica 1 раз на місяць перевіряла наявність документів Gopher й оновлювала свою базу даних ПОД'ів документів Gopher. У World Wide Web нічого подібного немає, для вирішення цього завдання використовують програми сканування мережі або роботи-індексатори. Розробка роботів – це досить нетривіальне завдання, оскільки існує небезпека зациклення робота або потрапляння на віртуальні сторінки. Усі пошукові системи мають свого робота, котрий проглядає мережу, знаходить нові ресурси, приписує до них терміни і поміщає в базу даних індексу. Головне питання полягає в тому, які терміни приписувати документам, звідки їх брати, адже багато ресурсів взагалі не є текстом. На сьогодні різні роботи використовують для індексування такі джерела для поповнення власних віртуальних словників: гіпертекстові посилання, заголовки

документів (title), заголовки абзаців (H1, H2 та інші), анотації, списки ключових слів і повні тексти документів, повідомлення адміністраторів про свої Web-сторінки. Для індексування Telnet, Gopher, ftp, нетекстової інформації використовують, як правило, URL, для новин Usenet і поштових списків – поля Subject і Keywords. Найбільший простір для побудови ПОДів дають HTML-документи. Проте не варто думати, що усі терміни з перерахованих вище елементів документів потрапляють у їхні пошукові образи. Дуже активно застосовують списки заборонених слів (stop-words), які не можа використати для індексування службових частин мови (прийменників, займенників тощо), а також часто проводиться нормалізація лексики. Отже, навіть те, що в пошуковій системі, наприклад, називається повнотекстовим індексуванням, насправді є вибиранням слів з тексту документа і порівнянням із низкою словників, після якого термін потрапляє у пошуковий образ документа, а потім і в індекс системи. Для того, щоб не надто розширювати словники й індекси, (а індекс Yandex, наприклад, становить більш ніж 62 ТВ (станом на початок 2009 р.), застосовують таке поняття як "вага" терміна – документ, зазвичай, індексується за допомогою 40–100 найбільш "важких" термінів.

Після того, як ресурси заіндексовані, тобто система склала сукупність пошукових образів документів, починається побудова пошукового апарату системи. Очевидно, що знайти власне потрібний файл або файли ПОДів забере багато часу, що абсолютно не прийнятно для інтерактивної системи Web. Для того, щоб швидко знайти інформацію в базі даних ПОДів, створюють індекси, структура і склад яких у кожній із систем відрізняються один від одного і залежать від багатьох чинників. До цих чинників належать і розмір масиву пошукових образів, й інформаційно-пошукова мова системи, і розміщення різних компонентів системи тощо. Розглянемо типову структуру індексу.

Він складається з таблиці ідентифікаторів сторінок (PAGE-ID), таблиці ключових слів (KEYWORD-ID), таблиці модифікації сторінок, таблиці заголовків, таблиці гіпертекстових зв'язків, інвертованого списку (IL) і прямого списку (FL).

PAGE-ID відображає ідентифікатори сторінок в URL цих сторінок; KEYWORD-ID – кожне ключове слово як унікальний

його ідентифікатор; таблиця заголовків призначена як ідентифікатор сторінки у її заголовках; таблиця гіпертекстових посилань – це ідентифікатор сторінок у гіпертекстове посилання на цю сторінку. Інвертований список подає до кожного ключового слова відповідний список пар (номер документа, ідентифікатор сторінки, позиція слова на сторінці), а прямий список – це сукупність пошукових образів сторінок. Усі ці файли так чи інакше використовують при пошуку, але головним серед них, безумовно, є файл інвертованого списку. Результат пошуку в цьому файлі – це об'єднання і/або перетин списків ідентифікаторів сторінок. Заключний список, який перетворюється на список заголовків, забезпечених гіпертекстовими посиланнями, повертається користувачеві у його браузер. Для того, щоб швидко шукати записи інвертованого списку, над ним надбудовують ще декілька файлів.

Для оновлення індексу застосовують комбінацію двох підходів: перший можна назвати корекцією індексу "на ходу", для цього слугує таблиця модифікації сторінок. Суть такого рішення досить проста: старий запис індексу посилається на новий, який і використовують при пошуку. Коли з'являється достатня кількість таких посилань для того, щоб помітити це при пошукові, то відбувається повне оновлення (або, як кажуть, перезавантаження) індексу.

Успіх інформаційно-пошукової системи з погляду швидкості пошуку визначається винятково архітектурою індексу. Як правило, спосіб організації цих обсягів є таємницею і гордістю компанії.

10.1.3. Інформаційно-пошукова мова системи

Індекс – це лише частина пошукового апарату, причому невидима для користувача. Другим його елементом є інформаційно-пошукова мова. ІПМ дає змогу сформулювати запит до системи в досить простій і наочній формі. Уже давно залишилася позаду романтика створення ІПМ як природної мови. Саме цей підхід використовували у системі Wais на перших стадіях її реалізації. Якщо навіть користувачеві пропонується вводити запити природною мовою, то це не означає, що система здійснюватиме семантичний розбір запиту

користувача. Як правило, фразу розбивають на слова, з цього списку видаляють заборонені та загальні слова, іноді проводиться нормалізація лексики, а потім їх усіх пов'язують або логічним AND, або OR.

Інший підхід полягає в обчисленні близькості між запитом і документом. Саме його використовують у Lycos, наприклад. Тоді, відповідно до векторної моделі представлення документів і запитів, обчислюється ступінь наближення. Сьогодні відомо дуже багато різних ступенів наближення. Найчастіше застосовують косинус кута між пошуковим образом документа і запитом користувача. Саме ці відсотки відповідності документа запиту і видають як довідкову інформацію до списку знайдених документів.

10.1.4. Типи інформаційно-пошукових мов

Головне завдання інформаційно-пошукової системи – це пошук інформації, релевантної потребам користувача. Слово "релевантність" означає відповідність між бажаною й отриманою інформацією. Релевантність можна ще пояснити як ступінь наближення між реально отриманими документами і тим, що хотіли отримати з системи. Природно, що тут виникає два завдання, які потрібно вирішити: представлення інформації в системі та формулювання інформаційних потреб користувача, які тісно пов'язані між собою.

Найбільш поширеними моделями представлення документів в інформаційно-пошуковій системі є різні варіації векторної моделі, коли документ подають як набір термінів. Як уже згадували раніше, це не весь текст документа, а тільки невеликий набір термінів, який відображає його зміст. Ґрунтуючись на такому уявленні про документ, і розглянемо різні інформаційно-пошукові мови.

10.1.5. Традиційні інформаційно-пошукові мови та їх модифікації

Найбільш поширеною ІПМ є мова, що дає змогу скласти логічні вирази з набору термінів. При цьому використовують булеві оператори AND, OR, NOT.

Така схема достатньо проста, і тому найширше застосовується в сучасних інформаційно-пошукових системах, хоча ще 20 років тому були добре відомі і її недоліки.

Булевий пошук погано масштабує видачу. Оператор AND може дуже скоротити кількість документів, які видаються на запит, при цьому практично все залежатиме від того, наскільки типовими для бази даних є пошукові терміни. Оператор OR навпаки – може привести до невиправдано широкого запиту, в якому потрібна інформація загубиться. Для успішного застосування цієї ІПМ потрібно добре знати лексику системи та її тематичну спрямованість. Як правило, для системи з такою ІПМ створюють спеціальні задокументовані лексичні бази даних зі складними словниками, які називають тезаурусами, вони містять інформацію про зв'язок термінів словника один з одним.

Модифікацією булевого пошуку є зважений булевий пошук, ідея якого достатньо проста: вважається, що термін описує зміст документа з певною точністю, і цю точність виражають у вигляді значущості терміна. При цьому зважувати можна і терміни документа, і терміни запиту. Запит можна формулювати за допомогою ІПМ, описаної вище, але видача документів при цьому ранжируватиметься залежно від ступеня близькості запиту і документа. При цьому визначення наближеності будується так, щоб звичайний булевий пошук був би окремим випадком зваженого булевого пошуку.

10.1.6. Мови типу "Like this"

При уважному розгляді зваженого пошуку закрадається природне бажання взагалі обійтися без логічних конекторів і визначати наближеність документа і запиту за допомогою інших критеріїв. Найбільш простою моделлю цього типу є лінійна модель індексування і пошуку, коли наближеність документа й запиту розглядають як кут між ними. Тоді обчислюють синус кута, який отримують як скалярний добуток двох векторів. Відповідно до значення ступеня наближення відбувається ранжирування документів при видачі посилань на них користувачеві. Скалярний добуток не дуже добре підходить для інформаційно-пошукових систем Інтернету, оскільки обсяг запиту, як правило, невеликий. Це у традиційних системах існували спеціальні служби, які коригували обсяг запиту, а в Інтернеті такі служби тільки з'являються. Тому реально застосовують інші способи наближення, але принцип залишається

той самий: спочатку обчислюють ступінь, а потім відбувається ранжирування.

Розглянутий підхід дає можливість більш конкретно розширити й уточнити запити, але він також не гарантує високих показників релевантності у разі вибору невдалої лексики.

10.1.7. Пошук в нечітких множинах

При цьому типі пошуку весь обсяг документів описують як набір нечіткої сукупності термінів. Кожен термін визначає певну монотонну функцію належності документам з їх загальної кількості. Коли подають запит AND, то це інтерпретується як мінімум із двох функцій, відповідних до термінів запитів, OR – як максимум, NOT – як 1 – <значення функції>. Залежно від набутих значень, результат пошуку також ранжирується, як і під час пошуку за способами наближення.

Потрібно відразу наголосити, що цей метод пошуку використовують лише в дослідницьких системах і він набув украй обмеженого поширення.

10.1.8. Порогові моделі

Як бачимо з попередніх розділів, на кінцевому етапі пошуку вибірка знайдених документів ранжирується. Але абсолютно очевидно, що способи наближення або пошук в нечітких множинах призводить до ранжирування усіх наявних у базі даних документів. Сучасні інформаційно-пошукові системи Інтернету мають бази даних тільки індексів, які мають обсяги у террабайтах. Ранжирувати такі обсяги неможливо, тому застосовують порогові моделі, що задають порогові значення для документів, які система шукає для користувача.

10.1.9. Кластерна модель і ймовірнісна модель інформаційного пошуку

У кластерній моделі можна використовувати два підходи. Перший полягає у тому, що весь обсяг заздалегідь розбивають на підмножини документів і під час пошуку обчислюють наближеність запиту до деякої підмножини. При іншому підході кластер “накручується” навколо запиту і найближчих до нього термінів. Найчастіше цю модель застосовують у системах, що уточнюють запит за релевантністю знайдених документів.

При використанні ймовірнісної моделі обчислюється ймовірність приналежності документа до класу релевантних запиту документів. При цьому використовують імовірність приналежності термінів запиту до кожного з документів бази даних.

10.1.10. Корекція запиту за релевантністю

Багато систем застосовують механізм корекції запиту за релевантністю. Це означає, що процедура пошуку має інтерактивний та ітеративний характер. Після проведення первинного пошуку користувач відзначає зі всього списку знайдених документів релевантні. На паралельні ітерації система розширює/уточнює запит користувача термінами з цих документів і знову здійснює пошук. Так продовжується доти, поки користувач не досягне порівняно найкращого результату. Корекція запиту за релевантністю – це досить поширений спосіб уточнення запитів. У деяких системах користувач може і не знати про те, що ця процедура застосовується. Наприклад, Google, у якому декілька ітерацій можуть виконуватися і без його втручання.

10.1.11. Інформаційно-пошукові мови Інтернету

Під час опису й класифікації інформаційно-пошукових систем ставилося завдання проаналізувати найбільш популярні і найбільш типові системи, які використовують у мережі.

Хотілося б звернути увагу читачів ще на один аспект вибору інформаційно-пошукової системи – це профіль її баз даних. Можна заперечити, що всі системи індексують одне і те ж – загальну кількість документів Інтернету, проте роблять вони це по-різному. Дуже важливий профіль системи, який задається розбиттям документів на теми і словники індексування, а також способом його підтримки. Як певний орієнтир тут можуть слугувати віртуальні бібліотеки.

10.1.12. Інтерфейс системи

Важливим чинником є вигляд представленої інформації в програмі-інтерфейсі. При цьому розрізняють два типи інтерфейсних сторінок: сторінки запитів і сторінки результатів пошуку.

При складанні запиту до системи використовують або меню-орієнтований підхід, або командний рядок. Перший підхід дає

змогу ввести список термінів, як правило, через пропуск і вибрати тип логічного зв'язку між ними. Логічний зв'язок розповсюджується на всі терміни. На поданій вище схемі (рис. **15**, с. 93) є так звані збережені запити користувача. У більшості систем це просто фраза на ІПМ, яку можна розширити завдяки додаванню нових термінів і логічних операторів. Але це тільки один спосіб використання збережених запитів. У традиційних системах його називають розширенням, або уточненням запиту, залежно від того, що отримуємо в результаті його перетворення: збільшення розміру вибірки або її скорочення. При цьому традиційна система зберігає не запит як такий, а результат пошуку, тобто список ідентифікаторів документів, який об'єднується/перетинається зі списком, отриманим під час пошуку документів за новими термінами.

Отже, результат пошуку в базі даних ІПС – це список покажчиків тих документів, що задовольняють запит. Більшість пошукових систем можуть подавати цей список у трьох варіантах залежно від налаштувань: тільки список посилань; список посилань із коротким описом; список посилань з розширеним описом. Розширений чи короткий опис запозичується або із заголовків, або зі змісту власне документа. Окрім цього система повідомляє, наскільки знайдений документ відповідає запиту. У Yahoo!!!, наприклад, повідомляється, скільки термінів із запиту міститься в пошуковому варіанті документа і відповідно до цього ранжирується результат пошуку. У Lycos визначається ступінь відповідності документа запитові й ранжирування проводиться за цим параметром. Звичайно, користувач має можливість уточнити запит.

Під час огляду інтерфейсів і засобів пошуку не можна залишити поза увагою процес коригування запитів за релевантістю. Нагадаємо, що релевантність – це ступінь відповідності знайденого системою документа потребі користувача. Розрізняють формальну релевантність та реальну. Формальна – це та, яку обчислює система і на підставі чого ранжирується вибірка знайдених документів. Реальна – це коли сам користувач оцінює знайдені документи. Деякі системи мають для цього спеціальне поле, де користувач може

відзначити документ як релевантний. При подальшій пошуковій ітерації запит розширюють термінами з цього документа, і знайдене знову ранжирується. Так відбувається доти, поки результат не стабілізується і не зупиниться, підібравши бажану вибірку.

10.2. Формулювання асоціативних запитів

Індуктивний метод пошуку інформації в Інтернеті можна реалізувати через побудову асоціативних запитів.

У загальному значенні під "асоціаціями" або "асоціативними запитами" мають на увазі список слів і словосполучень (запитів), який певною мірою пов'язаний з початковим запитом і формується більш-менш автоматично.

Деякі пошукові системи здатні формувати такий список. Наприклад, у Рамблері він з'явиться на сторінці результатів пошуку. У режимі проглядання знайдених документів за замовчуванням асоційовані запити розташовані в нижній частині цієї сторінки, після всіх знайдених документів. Вони розміщені під заголовком "У нас також шукають...". Це скорочений список запитів, тематично пов'язаних із початковим, тут можна перейти до іншого режиму перегляду знайденого, в якому доступний повний список асоціацій до початкового запиту, і його зручно проглядати, оскільки він розміщений в окремій зоні ліворуч.

Так можна порівняти свій початковий запит із тими формулюваннями і варіантами запитів, які використовували інші користувачі, і, за бажанням, вибрати варіант, що максимально уточнює тематику і сенс запиту. Наприклад, користувач набирає запит "реній" і отримує список сайтів з описами хімічної речовини, пропозиції купівлі/продажу, адреси сайтів, де можна стягнути наукову статтю або технічні умови застосування речовини тощо. Одночасно він показує список пов'язаних або асоціативних запитів, тобто має можливість вибору і миттєвого переходу до будь-якого елемента асоціативного ряду: метали, хімічні властивості, застосування, методи визначення і т. д.

10.2.1. Як формуються асоціації?

У загальному вигляді алгоритм автоматичного формування асоціацій побудований на такому припущенні: запити, які подав один і той самий користувач протягом деякого проміжку часу (одній пошуковій сесії), з великою ймовірністю мають одну і ту саму тематику. Якщо користувач, давши запит *x*, відразу після цього дає ще й запит *y*, можна припустити, що ці запити між собою пов'язані.

Якщо це припущення підтверджується внаслідок аналізу запитів від інших користувачів, тобто не один, а декілька відвідувачів сайту після запиту *x* переходять до запиту *y*, то всім новим користувачам, що подали запит *x*, можна порекомендувати пошукати ще і запит *y*. І навпаки, тим, хто подавав запит *y*, пропонується спробувати ще і запит *x*.

Імовірність тематичного збігу запитів *x* та *y* збільшується, якщо вони задані підряд, тому при автоматичному формуванні асоціацій враховуються саме пари "сусідніх" або розташованих один за одним запитів. Щодо проміжку часу, то за одиницю вимірювання беруть одну окремо взяту пошукову сесію, але якщо вона дуже велика, то часовий проміжок формування асоціацій обмежують. Цей проміжок підбирається експериментально і рідко становить понад 1,5 години.

Для формування асоціацій використовують протоколи роботи web-сервера пошукової системи. У них є дані про час звернення, мережеву адресу користувача (IP), унікальний ідентифікатор користувача і, звичайно ж, інформація про сам запит до пошукової системи. Унікальний ідентифікатор користувача – набір символів, який зберігається на його комп'ютері в ділянці так званих кук (cookies). Ідентифікатор формується, коли користувач вперше потрапляє на будь-яку web-сторінку пошукового сервера і відтоді міняється дуже рідко.

До автоматичної процедури формування асоціацій входить декілька етапів:

– побудова списків приблизних асоціацій;

– об'єднання списків і формування власне асоціацій;

– ранжирування асоціацій;

– фільтрація асоціацій;

– збереження остаточного списку асоціацій.

Таку процедуру потрібно проводити також при організації конкретного пошуку за асоціативним механізмом, а цей метод, як уже було зазначено, найбільш популярний.

Із погляду векторної алгебри вектори списків асоціацій запиту *x* і списку асоціацій запиту *y* близькі.

10.2.2. Для чого асоціації користувачам?

Уточнення або розширення запиту

Найбільш поширена сфера застосування асоціативних запитів – це уточнення і деталізація початкового запиту задля підвищення релевантності знайдених документів. Як наслідок, процедура пошуку інформації істотно спрощується і забирає менше часу.

Окрім того, простий перегляд списків асоціацій дає змогу виявити недоліки початкового формулювання запиту: його неоднозначність, можливість подвійного тлумачення, "розмитість". Як наслідок – відвідувач пошукового сайту має можливість на прикладі інших запитів від інших користувачів навчитися тому, як правильно ставити питання пошуковій системі, тобто, по-суті, скористатися "колективним розумом".

Наприклад, запит "диклофенак" приведе до появи величезної кількості знайдених документів, у якій перші декілька тисяч будуть цілком релевантними. Але знайти в цьому потоці сайтів потрібний – це непросте завдання. Недосвідчений користувач часто розгублюється, побачивши стільки інформації. А перегляд списку асоціацій: лікарська форма, рекомендації щодо застосування, хвороби, методи одержання, методи визначення, хімічні властивості та ін. – дає змогу зробити висновок, що слід звузити тематику запиту, тобто вказати, яка конкретно ділянка інформації потрібна.

Отримання нових знань на тему запиту

Часто буває так, що користувач уперше зацікавився певною темою. Наприклад, він хоче знайти інформацію про методики визначення певної речовини, але поки не знає, які хімічні та

фізичні властивості цієї речовини, які переваги тієї або іншої методики визначення тощо. Список асоціацій найчастіше містить не тільки уточнені й переформульовані запити, а й назви компаній, авторів методик, речовин, найбільш популяриних і поширених з цієї тематики. Наприклад, запит “диклофенак” приводить до появи асоціацій: мавалис, мелосикам, мідокалм, нимесил, найз, ібупрофен, вольт арен-гель та ін.

Як бачимо, автоматично сформований список містить не тільки найменування різних синонімів цієї речовини, але і назви популярних торгових марок тощо.

Виправлення помилок у написанні запиту

У деяких випадках асоціації надають користувачеві навіть таку можливість, як виправлення помилок у написанні запиту. В принципі, помилка при написанні запиту найчастіше призводить до відмови пошуку або до відсутності релевантних документів у списку знайденого. Наприклад, типова графічна помилка – використання латинської розкладки клавіатури для написання українського слова і навпаки. Оскільки ця помилка типова, то вона може відобразитися у списку асоціацій. Наприклад, запит “gjujlf” (“погода”, набрана латинською розкладкою клавіатури). Список асоціацій: прогноз погоди, погода в Москві, погода на тиждень, weather тощо.

Інший приклад – звичайні орфографічні помилки. Найскладніша ситуація для користувача виникає, коли він не в змозі цю помилку знайти. Як правило, тоді пошук виявляється невдалим, але деякі найбільш поширені помилки відображені в асоціаціях, і це може допомогти користувачеві. Наприклад, запит “фотоаПарат” асоціюється у пошуковій системі зі словом “фотоаППарат”, що відобразиться у списку пов’язаних запитів; або ж запит “Тольяті”, який приводить до появи асоціації “Тольятті” (правильне написання назви міста).

10.3. Бази даних

Бурхливий розвиток пошукових систем відбувався завдяки розвиткові технологій управління базами даних. Розвиток

електронних баз даних розпочав ще на початку 60 рр. XX ст. з розробок американської компанії “North American Rockwell”. З часом алгоритми обробки та збереження інформації значно вдосконалили, що забезпечило можливість швидко та ефективно оперувати значними обсягами різноманітної інформації.

Комп’ютерна база даних, або база даних (БД) – сукупність певним способом організованих даних, які зберігаються в пам’яті обчислювальної системи, та об’єктів, що відображають стан і взаємозв’язок цих даних.

Основне призначення БД полягає в тому, щоб одну і ту саму сукупність даних можна було використовувати для максимальної кількості різних програм.

База даних – це іменована сукупність даних і документів, що забезпечують ефективність зберігання та пошуку інформації. Водночас, база даних – це й електронне сховище інформації, доступ до якого здійснюється за допомогою одного або декількох комп’ютерів.

За структурою даних виділяють три класи БД: ієрархічні, мережеві, реляційні. Ієрархічні та мережеві моделі даних називають ще навігаційними.

Ієрархічна БД – взаємозв’язок елементів БД описується впорядкованим деревоподібним графом.

Прикладом ієрархічної БД є файлова система жорсткого диска комп’ютера: основою є жорсткий диск (корінь), від якого відходять товсті гілки (логічні диски), на котрих є дрібніші гілки (каталоги або директорії), а файли у цій моделі ніби листя. Так ми окреслюємо деревоподібну структуру розміщення інформації на комп’ютері.

Однією з переваг цієї моделі даних є наочність деревоподібних структур, проте зі збільшенням кількості інформаційних елементів (гілок та листя) у таких БД починають проявлятися й недоліки. Один із них, напевне, траплявся кожному, хто відкривав у своєму комп’ютері папку, що містить декілька тисяч навіть невеликого розміру файлів – тобто із розміром БД значно зростає період (і затрати обчислювальних ресурсів), потрібний для оперування такою БД. Інший недолік –

реальний світ далеко не завжди може з'являтися у вигляді деревоподібної структури з єдиним кореневим сегментом.

Ширші можливості надає мережева модель БД, що є більш загальним варіантом ієрархічної моделі (ієрархічна модель є частковим варіантом мережевої).

Мережева БД – взаємозв'язок елементів БД, який описується довільним графом.

Прикладом мережевої моделі БД є Інтернет. Величезна кількість окремих інформаційних елементів (веб-сторінок) пов'язані між собою за допомогою гіперзв'язків. Аналогом мережевої БД є також головний мозок, у якому нейрони за допомогою синапсів пов'язані з іншими нейронами.

Головною перевагою мережевої моделі БД над ієрархічною є швидша навігація. Щоб в ієрархічній БД перейти на віддалену гілку, потрібно спуститися до більш товстої гілки, що є спільною для обох маленьких гілочок, а потім уже піднятися. У мережевій БД можна зразу ж переміститися з однієї маленької гілочки на більш віддалену.

Недоліки мережевої моделі даних є зворотною стороною її переваг – кількість можливих зв'язків між елементами зростає пропорційно до квадрата кількості елементів. А оскільки далеко не завжди існують безпосередні зв'язки між елементами, то така схема БД стає досить складною для розуміння й ефективної обробки інформації. Крім того, в мережевій моделі внаслідок можливості встановлення довільних зв'язків між елементами, значно послаблюється контроль цілісності власне БД, що ми, як правило, і спостерігаємо на прикладі Інтернету.

З огляду на потребу в ефективній обробці великого обсягу інформації у 1970 рр. XX ст. Едгар Ф. Кодд (Edgar F. Codd) розробив реляційну модель БД, котра спочатку викликала зацікавленість лише у наукових колах. Масового застосування БД на цій основі набули тільки упродовж останніх 10–15 років.

Реляційна БД – взаємозв'язок елементів, що описується комплексом взаємопов'язаних таблиць.

Ілюстрацією реляційної бази даних може бути багатовимірна таблиця. Завдяки одній її властивості – кожен елемент інформації в базі трапляється лише один раз і

розташований у цій таблиці на перетині відповідних гіперрядків та гіперстовпчиків. Людина нормально сприймає двовимірні таблиці, котрі по-суті є двовимірними проекціями всієї бази даних. Оскільки таку базу даних можна розділити на декілька взаємопов'язаних таблиць, то за цю властивість вона отримала назву реляційної (англ. Relation – відношення).

Реляційна модель даних забезпечує низку важливих можливостей, які сприяють управлінню БД, а їх використання робить порівняно легким, стійким до помилок і передбачуваним. Основна перевага реляційних баз даних – швидкий пошук у базі даних. Сучасні бази даних можуть містити мільярди елементів, проте пошук забирає менше секунди. У цьому можна переконатися, посилаючи запит на пошуковий сервер – результат отримуємо майже миттєво. Бази даних пошукових систем реляційні.

У реляційній базі даних інформація міститься у вигляді таблиць.

10.4. Питання для самоперевірки

1. Для чого застосовують пошукові системи?
2. Яка структура інформаційно-пошукових систем?
3. Поняття пошукового образу документа.
4. Що таке індексна база пошукової системи?
5. Інвертований список – це …
6. Назвіть основні типи інформаційно-пошукових мов.
7. Поняття інтерфейсу системи.
8. Що є результатом пошуку в базах даних інформаційно-пошукових систем?
9. Як побудовані асоціативні запити?
10. Для чого потрібне дослідження автоматично побудованих асоціацій?
11. Що таке база даних?
12. Назвіть основні типи баз даних.
13. Які переваги та недоліки різних типів баз даних?
14. На основі знань про бази даних охарактеризуйте потребу у створенні пошукових систем.
15. Які типи баз даних використовують в інформаційно-пошукових системах?

11. ПРИНЦИПИ ФОРМУВАННЯ ПОШУКОВИХ БАЗ ДАНИХ

Для того, щоб ефективно здійснювати пошук інформації в Інтернеті, потрібно розуміти основні принципи побудови пошукових баз даних.

11.1.1. Як влаштована пошукова система?

Метою пошукових систем, як відомо, є надання матеріалу, котрий максимально відповідає введеному користувачем запиту. Існує безліч варіантів алгоритмів. У цьому розділі ми торкнемося лише тих, які безпосередньо стосуються веб-майстрів та оптимізаторів, а саме: алгоритмів популярних пошукових систем типу Yandex, Google, Rambler та інших.

Класичні пошукові системи, такі як описані вище, складаються з двох частин, що працюють паралельно одна з одною:

– індексатор – система, що стягує документи (HTML файли, зображення та ін.), які цікавлять користувача, та розбиває їх на слова, лексеми, токени, придатні для зберігання в базі даних;

– пошукова система – набір програмного забезпечення, який надає доступ до інформації, що зберігається, відповідно до запиту користувача.

11.2. Отримання інформації з мережі

Для отримання інформації задля подальшого індексування використовують програми, що називаються “павуками” (з англ. Spider), роботами (з англ. Crawler), чи ботами (з англ. Bot).

Завдання будь-якого “павука” – опрацювати якомога більше документів у мережі протягом вказаного часу, забезпечивши повноту пошуку (кількість документів) і його актуальність (новизна документів, що зберігаються в індексі пошукача). Перехід від одного документа до іншого відбувається за допомогою посилань, які додаються до списку документів, що підлягають опрацюванню.

Як правило, пошукова система запускає одночасно тисячі таких модулів, кожен із яких сканує свій сегмент, обмежений доменом або набором документів, які підлягають почерговому скануванню. Чекати стягування документів почергово немає сенсу, оскільки кожен із них не залежить від іншого. У всіх

“павуків” є диспетчер, що регулює їх звернення до сайтів і розподіляє навантаження на опитувані сервери, щоб запобігти їх перенавантаженню.

Для стягування “павуки” користуються таким самим протоколом отримання інформації з сервера, як і звичайні браузери, – протоколом HTTP. Тому вони мають “бачити” те саме, що і звичайний користувач, за винятком непотрібної йому інформації (наприклад, картинки у текстовій пошуковій системі). Надання “павукові” й користувачеві різної інформації називається клоакінгом (cloaking), це вважається порушенням правил і на сьогодні карається власниками пошукових систем виключенням з індекса (відсутністю сайту в індексному файлі, за яким відбувається пошук і, як наслідок, відсутністю сайту в результатах пошуку).

Якщо в певний момент документ не можна стягнути, то “павук” робить повторну спробу. Після деякої кількості невдалих спроб документ видаляється з індексного файла і, як наслідок, з результатів пошуку, якщо він там містився.

Стягування можна організувати за різними принципами: завширшки, завглибшки, за PageRank або індексом цитованості, проте його завдання залишається єдиним – звести до мінімуму трафік, отримавши максимум інформації. Роботам стягування інформації активно допомагають конструкції, що керують (або підказують) на кожному сайті: файли robots.txt (які визначають, що не дозволено стягувати), мета-теги nofollow, noindex (вказують, чи варто переходити за посиланнями в певному документі або зберігати його), а також такі розробки, які рекомендують роботові стягувати особливо актуальні сторінки.

11.2.1. Зберігання інформації

Більшість доступних у мережі пошукових машин побудовані на основі інвертованого індексу – стягувані документи розділяють на слова, а потім за словами складають таблицю відповідностей слів і документів, які їх містять. Класичним прикладом інвертованого індексу є глосарій термінів у кінці книги, що вказує, на якій сторінці (або у якому документі в термінах пошукової системи) трапляється вказаний термін. Робота з таким типом індексів набагато ефективніша порівняно,

наприклад, з тим же прямим пошуком: у разі інвертованого індексу ми беремо номери сторінок, на яких міститься слово, і видаємо користувачеві відповідні сторінки, а під час прямого пошуку ми змушені були б перегорнути книгу наново, шукаючи відповідні місця.

При індексації, насамперед, документ набуває текстового вигляду з використанням різних модулів обробки форматів або парсерів (від англ. Parser – аналізатор). Сучасні системи шукають й індексують не тільки текстові й html-документи, а й працюють із низкою інших форматів, наприклад, pdf, doc, swf тощо. Парсер має розпізнати формат документа і виділити зміст, зберігши його структуру: інформацію про заголовки, списки, речення, визначення та інші елементи розмітки, що стосуються не оформлення тексту, а його структури. Структура документа використовується для обліку важливості слів у ньому при пошукові (term weight): слова, що містяться в заголовку самого документа, важливіші, ніж ті, що розташовані в заголовках параграфів. Своєю чергою слова у заголовках у тексті документа (наприклад, заголовки h1–h6 у разі стандарту html) мають важливіше значення під час пошуку, ніж слова з параграфів або нумерованих списків. Відповідно до структури кожному слову надається ступінь вагомості у документі.

Документи, що складаються з речень, розділяють на слова (токени). Найчастіше поділ відбувається за пропусками, переходами рядків й іншими символами, що є ознаками розриву слів. Іноді до ознак розриву належить зміна регістра літер або зміна буквено-числових послідовностей – тоді слово “CanonEos888” буде розділено на три: “Canon”, “Eos” і “888”. Розділеним словам надають уніфікованого вигляду (нормалізують), наприклад за відмінками, числами або родами. Так, однакові слова, що вказані у різних формах подають, наприклад, у називному відмінку однини. По-перше, це дає змогу точніше працювати з подальшим пошуком документів: визначивши кількість різних форм одного і того самого слова (частоту вживання терміна) в потрібному документі, ми можемо точніше вказати на документи, відповідні (релевантні) до пошукового запиту, сортуючи (ранжируючи) їх за ступенем відповідності. По-друге, використовуючи механізми нормалізації слів, можна

видати користувачеві результати його запиту, що містять форму слова, яка відрізняється від форми цього самого слова в іншому документі: наприклад, у документі може міститися словосполучення “хорошої машини”, але він має з’явитися користувачеві на запит “хороша машина”.

Безумовно, кінцевий (розділений на токени) текст, призначений для обмеженого застосування, може виглядати достатньо кумедно: наприклад, речення “Собака біг по полю, де було багато левів” перетвориться в “Собака бігти поле/полоти є багато лев/Львів”, проте його “комічність” найкраще підходить для зберігання інформації та подальшого пошуку за її допомогою. Треба зазначити, що у разі використання спрощених алгоритмів, які не спираються на семантику тексту, не завжди вдається точно встановити частину мови, що часто призводить до подвійності та неоднозначності деяких термінів (дизамбігурації).

Використовуючи алгоритм інвертованого індексу, виникає багато обмежень для файла індексу, щоб не спричинити надмірного його розростання. Наприклад, наявність частки “і” або прийменника “на”, які вживаються достатньо часто в кожному документі або навіть реченні, призвів би до того, що таблиця відповідності цих слів і сторінок, на яких розташовані вони, містила б мільйони значень. Це істотно зменшило б обсяг бази, ніяк не вплинувши на його якість (пошук подібних слів украй рідко затребуваний на практиці).

Для того, щоб уникнути подібних ситуацій, на етапі індексування вводять обмеження у вигляді стоп-слів, які індексатор ігнорує. Щоб не перевантажувати індексний файл зайвими словами, в подібні списки потрапляють, в основному, службові слова: прийменники, займенники, частки, артиклі й інші лексеми, що рідко використовують при пошукові, але які часто вживаються у документах. Як приклад, файл стоп-слів російського пошукового Yandex Server містить такі слова, як “де”, “який”, “ваш” та ін. Ці самі слова обрізають і не враховують під час запиту у базі даних, оскільки вони не потрапляють туди при індексації.

Отриманий набір слів перевіряє значна кількість алгоритмів на наявність дублів у базі. У Yandex використовується алгоритм шинглів, про які йтиметься в подальших розділах. Проста перевірка полягає в підрахунку контрольних сум усіх слів

тексту або частини слів, обмежених певною ознакою (наприклад, при розрахунку контрольної суми беруть до уваги тільки іменники). Документи, визнані дублями, або не беруть до уваги під час пошуку, або розташовані лише в кінці результату пошуку.

11.3. Питання для самоперевірки

1. Вкажіть основні компоненти пошукової системи.
2. Яку функцію виконують пошукові роботи?
3. Які елементи веб-сторінок призначені для пошукових роботів?
4. Що таке парсер?
5. Що таке нормалізація слів?
6. Що таке стоп-слова?

12. ПОШУКОВІ ЗАПИТИ

12.1. Мова пошукових запитів

Коли ми шукаємо в Інтернеті певну інформацію, то, як правило, використовуємо пошукові системи і сподіваємося витратити якомога менше часу на цей процес. Давайте розглянемо в загальних рисах процес роботи пошукових систем, а саме – видачі результатів пошуку.

Пошукові системи постійно удосконалюють свої механізми, але і користувачеві не зашкодить уміння правильно скласти запит для того, щоб отримати точну відповідь на нього. Запит має бути конкретизований, щоб пошукова система могла знайти найбільш відповідний результат. Наприклад, слово “диски”, може мати декілька значень: диски автомобільних коліс, компакт-диски, диски для метання, а можна мати на увазі диски якоїсь планети. Тому, якщо ми шукаємо, наприклад, компакт-диски, то і запит потрібно складати відповідно.

Передусім, під час пошуку інформації в Інтернеті важливо мати на увазі, що ступінь релевантності з погляду пошукового механізму тим вищий, чим ближче розміщені шукані слова одне до одного у виявленій фразі та чим менше між ними вказано інших слів.

Строгий пошук, тобто строгу відповідність результатів

пошуку шуканому слову або фразі можна отримати, застосовуючи так звану мову пошукових запитів. Правильне застосування мови пошукових запитів істотно покращує його точність, тому для якісного пошуку інформації в мережі варто знати і застосовувати прийоми складання запитів. Запит може складатися з одного або декількох слів, може містити розділові знаки, цифри.

Мова пошукових запитів – це набір службових команд – символів та операторів, які дають можливість деталізувати запит, що посилається пошуковій системі. У ньому використовують прості логічні функції об'єднання або усування слів із запиту. Слід нагадати, що службові слова в різних пошукових системах можуть відрізнятися, але є певний усереднений набір, уживаний практично у всіх пошукових систем.

12.1.1. Логіка пошукових запитів

Якщо пошук інформації малопродуктивний, для підвищення його точності можна застосувати логічні оператори. Їхнє використання підтримують майже всі пошукові системи.

Існує три основні логічні оператори: логічне "І" (AND), логічне "АБО" (OR) та логічне "НІ" (NOT).

Давайте розглянемо роботу пошукових систем та логічних операторів на прикладі модельної бази даних. Нехай ця база даних описує вісім різних документів, що містять сім різних ключових слів. Вона буде подавати документи в дуже спрощеному вигляді – якщо слово є в документі, йому відповідає символ "+" у базі даних (реально це має бути коефіцієнт значущості слова в документі), якщо немає – будемо писати пропуск, оскільки в реальній базі даних відповідних клітинок взагалі немає – вони з'являються тільки тоді, коли ми подаємо базу даних як таблицю.

Таблиця 1

Модельна база даних

Слова	**Документи**							
	1	2	3	4	5	6	7	8
Літій	+	+	+	–	–	–	–	–
Натрій	+	–	–	+	–	+	–	+

Калій	+	+	–	–	+	–	+	–
Магній	+	–	–	–	–	–	–	+
Реній	–	–	+	–	–	–	+	–
Осмій	–	–	–	–	+	–	–	–
Цезій	+	–	–	–	–	+	+	–

12.1.1.1. Простий запит

Простий запит за одним ключовим словом видасть тільки ті документи, котрі його містять.

Приклад 20. *Простий запит:*

На запит "Літій" система видасть документи 1, 2, 3; на запит "Реній" – документи 3 і 7.

12.1.1.2. Логічний оператор AND

Два слова, сполучені оператором AND (логічне І) утворюють складний запит, який задовольняють тільки ті документи, які одночасно дають результат для обох цих запитів. Це означає, що шуканий документ обов'язково має містити обидва слова.

Приклад 21. *Логічний оператор AND:*

На запит "Літій AND Калій" система видасть документи 1 і 2; на запит "Натрій AND Цезій" – 1 і 6.

Використання оператора AND значно зменшує обсяги пошуку.

12.1.1.3. Логічний оператор OR

Два слова, сполучені оператором OR (логічне АБО), утворюють складний запит, який задовольняють тільки ті документи, які містять будь-яке зі слів запиту.

Приклад 22. *Логічний оператор OR :*

На запит "Літій OR Калій" система видасть документи 1, 2, 3, 5 і 7; на запит "Магній AND Осмій" – 1, 5 і 6.

Використання оператора OR збільшує обсяги пошуку.

12.1.1.4. Логічний оператор NOT

Якщо перед словом розміщений оператор NOT, це означає, що його потрібно вилучити із результатів пошуку. Документи, які його містять, не відобразяться у списку знайдених.

Приклад 23. *Логічний оператор NOT:*

На запит "Літій NOT Калій" система віднайде документи 2 і 3; на запит "Натрій NOT Цезій" – 4 і 8.

Подібно до оператора AND, оператор NOT зменшує обсяги пошуку. Проте його варто використовувати за наявності зайвої інформації у результатах пошуку.

12.1.1.5. Пріоритет логічних операторів

Якщо у запиті наявні декілька логічних операторів, вони мають певну ієрархію під час виконання. Спочатку береться до уваги оператор NOT, потім – AND, а тоді – OR.

Приклад 24. *Пріоритет логічних операторів:*

На запит "Літій AND Калій OR Магній" будуть видані документи 1, 2 і 8; на запит "Літій OR Калій AND Магній" – будуть видані документи 1, 2 і 3; результатом запиту "Літій OR Калій NOT Магній" буде знайдено документи 2, 3, 5 і 7.

12.1.1.6. Оператори за замовчуванням

Якщо оператор явно не вказаний (ключові слова записані підряд), то пошукові машини за замовчуванням здебільшого використовують оператор AND, знаходячи тільки документи, що містять усі слова запиту.

Приклад 25. *Інтерпретація запиту пошуковою машиною:*

Запит "Літій Калій" машина перетворить у запит "Літій AND Калій" і видасть документи 1 і 2.

12.1.1.7. Альтернативні записи логічних операторів

Оскільки записувати логічні оператори у запиті незручно, то в більшості пошукових систем використовують альтернативний запис. Потрібно зазначити, що використання альтернативних записів логічних операторів у зв'язку з різними алгоритмами розрахунку релевантності може призводити до переоцінки або недооцінки пошукових термінів, тому ці оператори бажано використовувати у звичайному вигляді.

12.1.1.8. Інші загальноприйняті оператори

Часто замість логічних операторів використовують їх альтернативи. Як уже було сказано, вони діють не зовсім так, як вихідні оператори. Вони складаються з одного–двох знаків та записуються безпосередньо перед ключовим словом.

12.1.1.9. Оператор "+"

Оператор "+" виконує роль підсиленого AND; пошуковий термін з цим оператором у запиті набуває більшої значущості. Це можна стверджувати з огляду на кількість знайдених під час пошуку документів.

Приклад 26.

Запит "натрій AND калій":
www.google.com – *335 000 сторінок;*
www.yandex.ru – *31 000 сторінок.*
Запит "натрій + калій":
www.google.com – *13 800 сторінок;*
www.yandex.ru – *31 000 сторінок.*

12.1.1.10. Оператор "-"

Оператор "-" виконує роль послабленого NOT, пошуковий термін з цим оператором у запиті набуває меншої значущості, на що вказує кількість знайдених під час пошуку сторінок.

Приклад 27.

Запит "натрій NOT калій":
www.google.com – *43 700 сторінок;*

www.yandex.ru – *31 000 сторінок.*
Запит "калій –натрій":
www.google.com – *3 770 000 сторінок;*
www.yandex.ru – *3 000 000 сторінок.*

12.1.1.11. Оператор "|"

Оператор "|" – як посилений OR, пошуковий термін з цим оператором трактується у більш широкому сенсі.

Приклад 28

Запит "натрій OR калій":
www.google.com – *1 260 000 сторінок;*
www.yandex.ru – *31 000 сторінок.*
Запит "калій |натрій":
www.google.com – *1 260 000 сторінок;*
www.yandex.ru – *171 000 сторінок.*

Так, запит "інформація технології кредит" пошукова система тлумачитиме як "інформація AND технології AND кредит". На сторінці розширеного пошуку оператора за замовчуванням можна замінити на OR (шукати слова запиту: хоч би одне).

Оператори AND і OR мають скорочені позначення:

AND – &;

OR – |.

Запит із декількох слів із зазначеними операторами система тлумачитиме відповідно до їх пріоритету: оператори AND і NOT традиційно мають вищий пріоритет, тому запит з декількох слів при обробці спочатку групується за операторами AND і NOT, і лише потім – за оператором OR. Змінити порядок групування можна за допомогою дужок.

Оператор "OR" (логічне АБО) і його скорочення "|" дає змогу шукати хоч би одне зі слів ключової фрази. Наприклад, запит "автомобіль OR Toyota", "автомобіль | Toyota" дасть в результаті хоча б одне слово з цього поєднання. Хоча в цьому разі він, найімовірніше, знайде обидва слова, оскільки вибраний невдалий приклад.

Наступний символ – "" (подвійні лапки), використовується для пошуку цілої фрази. Для того, щоб спеціальні символи були

розцінені як звичайні, фразу слід подати у подвійних лапках.

Подвоєння певної службової команди означає, що ключове слово має бути не тільки в реченні, а й у всьому документі. Наприклад, фраза “концентрована нітратна кислота ~~ціна ~~продаж ~~виробництво” означає, що слів “ціна”, “продаж”, “виробництво” не має бути в межах усього документа, а не тільки в реченні.

Крім того, існує ще низка дуже корисних службових специфічних символів, які дають можливість шукати й отримувати важливі дані. Синтаксис може відрізнятися для різних систем, але в цілому він схожий.

Таблиця 2

Оператори пошукових запитів

Синтаксис	**Функція оператора**	**Приклад**
$title (вираз)	пошук у заголовку	$title (CompTek)
$anchor (вираз)	пошук у тексті посилань	$anchor (CompTek | Dialogic)
#keywords =(вираз)	пошук у ключових словах	#keywords=(пошукова система)
#abstract =(вираз)	пошук в описі	#abstract= (пошук)
#image =”значення”	пошук файла зображення	#image=”tort”
#hint =(вираз)	пошук у підписах до зображень	#hint=(lenin | Ленін)
#url =”значення”	пошук на заданому сайті (сторінці)	#url=”www.comptek.ru”
#link	пошук посилань на	#link=”www.yandex.ru”

=”значення”	заданий URL	
#mime =”значення”	пошук у документах певного (pdf або rtf) типу	#mime=”pdf”

Усі приклади показані у загальних рисах, і кожна пошукова система має для уточнення пошуку власний набір символів-операторів, і тому практично в кожній пошуковій системі про це існує докладна довідка. Вивчивши такий набір символів і прийоми їх застосування, можна зробити свій пошук в Інтернеті значно ефективнішим і швидким.

12.2. Особливості оформлення пошукових запитів

Пошуковий запит може складатися з одного або декількох слів, включно з розділовими знаками. Складати прості запити можна і не вдаючись до тонкощів мови запитів. Так, якщо ввести в пошуковий рядок декілька слів без розділового і логічних операторів знаків, будуть знайдені документи, що містять усі ці слова (причому на обмеженій відстані один від одного).

Проте знання і правильне застосування мови запитів пошукової системи допоможе зробити Ваш пошук швидким та ефективним.

12.2.1. Регістр

Загалом, регістр написання пошукових слів й операторів значення не має, тобто будинок і БУДИНОК, Not і not сприймаються однаково. І лише іноді, задля підвищення якості пошуку, регістр слів пошукового запиту береться до уваги.

12.2.2. Оператори

Запит, що складається з декількох слів, може містити й пошукові оператори. Пошук операторів у документі не проводиться, вони слугують лише інструкцією для пошукової системи. Усі оператори пошукової системи бінарні, тобто мають ліву і праву частини, кожна з яких також є запитом (що за

замовчуванням складається з одного слова). Для зміни сфери дії операторів (сукупність декількох слів запиту в аргумент оператора) застосовують дужки і лапки.

12.2.3. Лапки

Для пошуку цитат можна використовувати подвійні лапки. Слова запиту, взяті у подвійні лапки, шукають у документах саме в тому порядку й у тих формах, у яких вони подані в запиті.

12.2.4. Дужки

При формуванні запитів іноді виникає потреба об'єднати слова запиту в групи, які будуть аргументами для деякого оператора. Такі групи подають у дужках.

Частина запиту, взята в дужки, власне і є запитом, на неї розповсюджуються правила мови формування запитів. Використання дужок дає змогу складати вкладені запити і передавати їх операторам як аргументи, а також перекривати пріоритети операторів, прийняті за замовчуванням.

Приклад 29. *Складний запит:*

Запит "натрій AND (калій OR магній)" еквівалентний запитові "натрій AND калій OR натрій AND магній" (не забуваємо про порядок надання переваги операторам).

12.2.5. Метасимволи

Пошукові системи, як правило, не підтримують запити з використанням метасимволів ("*", "?"), які на персональному комп'ютері використовують у значенні "будь-який підрядок" і "довільний одиничний символ" відповідно. Можливість використання метасимволів треба з'ясувати у довідкових системах відповідних пошукових машин. Найчастіше метасимвол "*" використовують у значенні "будь-яке слово".

12.2.6. Застосування мови запитів

Кожен запит, адресований пошуковій системі, обробляється відповідно до правил мови запитів. Деякі слова та символи трактуються як оператори мови запитів і спеціально обробляються. Фактично мовою запитів описується певна

формула, яку використовують при пошукові, – кожен із документів "зіставляється" з нею, і результатом пошуку є тільки ті документи, які її задовольняють.

12.2.7. Морфологія

Сучасні пошукові системи зараз враховують правила словозміни відповідної мови. Звичайно, вітчизняні пошукові системи мають більш розвинутий морфологічний апарат. Нещодавно морфологію української та російської мов почали вирізняти також у Google.

Приклад 30. *Врахування морфології слів:*

На запит "металевий" система видасть документи, що містять слова: металевого, металеві, металевим та ін.

12.2.8. Стоп-слова

Деякі слова і символи, за заявами пошукових сервісів, видаляються із запиту внаслідок низького рівня інформативності. Це так звані стоп-слова – найбільш поширені слова української, російської й англійської мов, наприклад, прийменники, частки й артиклі. Їхні значення можуть уповільнити пошук і негативно вплинути на якість і зміст результатів. Проте практика засвідчує, що службові слова певною мірою індексуються й існують у базах пошукових сервісів, і якщо задати запиту службове слово, можна одержати значний список документів. Можна припустити, що службові слова видаляються з багатослівних запитів, хоча перевірка дає неочікувані результати.

Приклад 31. *Перевірка на стоп-слова Yandex:*

Подайте пошуковій системі Yandex запити "реній" та "і реній" (без лапок), оцініть результат.

Втім, коли потрібно зазначити ці слова у запиті, треба взяти словосполучення в лапки або знайти точну фразу у розширеному пошукові.

12.2.9. Обмеження відстані

Якщо до запиту входить одне або декілька слів без застосування

операторів і конструкцій мови запитів, то будуть знайдені документи, які містять усі слова запиту. При цьому для кожного запиту завжди існує так зване обмеження контексту – позитивне число, яке за замовчуванням дорівнює відстані у 40 слів. Документ, у якому вжиті усі слова запиту, з'явиться тільки тоді, якщо відстань між словами запиту буде меншою від цього числа. Наприклад, за запитом "діамантовий зелений" система знайде ті документи, у яких слова "діамантовий" і "зелений" хоча б один раз вживатимуться менше, ніж через 40 слів один від одного.

Значення обмеження контексту можна змінювати конструкцією "число, запит", де число – будь-яке позитивне число, запит – будь-який коректний з погляду пошукової системи запит, що складається більш ніж з одного слова (очевидно, обмеження відстані між словами у разі однослівного запиту не має сенсу). Отже, за запитом "2, діамантовий зелений" буде знайдено тільки ті документи, у яких між словами "діамантовий" і "зелений" хоча б раз не повторюється жодного слова (оскільки лише у разі їх безпосереднього сусідства різниця в порядкових номерах слів не перевищує 2, тобто дорівнює 1).

12.2.10. Незнайдені слова

Якщо запит складається з декількох слів, і при цьому деякі з них взагалі відсутні в індексі пошукової системи, то ці слова ігноруються – результатом пошуку буде частковий запит, у якому немає видалених в індексі слів. При цьому на сторінці результатів пошуку видається відповідна діагностика.

12.3. Питання для самоперевірки

1. Що таке мова пошукових запитів?
2. Назвіть основні логічні оператори.
3. Поясніть механізм дії логічного оператора "І".
4. Характеризуйте механізм дії логічного оператора "АБО".
5. Поясніть вживання логічного оператора "НЕ".
6. Укажіть послідовність пріоритету виконання логічних операторів?
7. Який логічний оператор пошукова система використовує за замовчуванням?
8. Назвіть інші оператори, які в запитах використовують

пошукові системи.
9. Як працює оператор “+”?
10. Призначення оператора “-”?
11. Для чого застосовують оператор “|”?
12. Наскільки важливий регістр символів у пошуковому запиті?
13. Яку функцію у запиті виконують подвійні лапки?
14. Для чого у запиті використовують дужки?
15. Чи можна використовувати метасимволи у запитах?
16. Які пошукові системи беруть до уваги морфологію?
17. Яким буде результат пошуку, якщо пошукова система не “розпізнає” певне слово у запиті?

13. МЕТОДИ ПОШУКУ ІНФОРМАЦІЇ

До категорії “інформація” належить і поняття “бібліографічна інформація”, яка є одним із найважливіших з її складників. Це пояснюється тією функціональною роллю і соціальним завданням, які вирішує суспільство за допомогою бібліографічної інформації в процесі виробництва, розповсюдження і використання соціальної інформації, знань. Наука, що вивчає історію, теорію і методику інформаційного пошуку називається бібліографічною евристикою.

Об’єктивний характер існування та використання такої інформації і зумовлює найрізноманітніші завдання її пошуку. В цілому всі ці завдання можна звести до трьох основних напрямів інформаційного пошуку, які визначають і три основні його види: бібліографічний, документальний і фактографічний.

Бібліографічний – пошук потрібних відомостей про певне джерело і встановлення його наявності у системі інших джерел.

Здійснюється через розшук бібліографічної інформації та бібліографічних посібників (інформаційних видань), що спеціально створюється для ефективнішого пошуку і використання інформації (літератури, книги);

Документальний – пошук власне інформаційних джерел (документів і видань), у яких є (або може міститися) потрібна інформація.

Фактографічний – пошук фактичних відомостей, що містяться в літературі, про історичні факти і події; технічні характеристики машин і процесів; властивості речовин і матеріалів; біографічні дані з життя і діяльності письменника, ученого та ін.

Аналітичний – пошук аналітичної інформації, що ґрунтується на дослідженнях різного масштабу.

Наприклад, щоб знайти інформаційне джерело (документ або видання), треба знати певну сукупність бібліографічної інформації (фактів), що характеризує його, відрізняє від багатьох інших: наприклад, від написаних тим же автором, на аналітичну тему, випущених у тому самому видавництві, у певний рік тощо. І навпаки: щоб здійснити фактографічний пошук у певній галузі знань або практики, потрібно спочатку знайти ті літературні джерела (документи, видання), у яких можуть міститися факти, що нас цікавлять. Тому спершу треба провести бібліографічний і документальний пошук. У будь-якому разі потрібно спочатку виробити стратегію пошуку.

Стратегія пошуку – загальний план (концепція) дій системи або користувача для формування і задоволення інформаційної потреби користувача, зумовлений особливостями його мети і видом пошуку, так і системними "стратегічними" рішеннями – архітектурою бази даних, методами і засобами пошуку в конкретній пошуковій системі.

Вибір стратегії загалом є оптимізаційним завданням. Це означає, що сюди належить декілька кроків, й успішних, і неуспішних, котрі допомагають досягти бажаної мети. На практиці значною мірою він визначається можливістю досягнення компромісу між практичними потребами і можливостями наявних засобів.

Метод пошуку – сукупність моделей і алгоритмів реалізації окремих технологічних етапів: побудови пошукового образу запиту (ПОЗ), відбору документів (зіставлення пошукових образів запитів і документів), розширення і реформулювання запиту, локалізації та оцінки видачі.

Для цього бібліографічна евристика має низку методів пошуку інформації: суцільний, вибірковий, інтуїтивний, типологічний (рецептурний), індуктивний, дедуктивний, метод бібліографічних посилань, метод сходження від абстрактного до конкретного.

Коротко їх розглянемо.

13.1. Загальні методи пошуку інформації

Суцільний метод – це обстеження усіх наявних джерел інформації.

Очевидно, що в сучасних умовах цей метод важко реалізовується навіть для вузької теми, тому часто його розглядають лише як теоретичний.

Вибірковий метод – раціональніший і реальніший спосіб пошуку інформації, який зводиться до підбору і вивчення певної сукупності джерел.

Інтуїтивний метод – уміння знайти потрібний матеріал за допомогою бібліографічного чуття, розвинутого завдяки тривалій практиці.

Типологічний (рецептурний) метод – використання завдань інформаційного пошуку дає змогу для кожного з них використовувати певний спосіб і документальні засоби (посібники, джерела) вирішення, тобто дати своєрідну модель, щоб знайти потрібні матеріали.

Так, оптимальною типовою моделлю для розробки різних “рецептів” інформаційного пошуку є бібліографічний опис.

Індуктивний метод інформаційного пошуку – процес пізнання з використанням індукції здійснюється від одиничних думок і фактів до загальних правил, у яких виражена загальна закономірність.

Дедуктивний метод інформаційного пошуку – це зворотний щодо індуктивного методу спосіб, тобто процес пошуку відбувається від загального до конкретного.

Таку спрямованість думки можна застосувати під час дослідження певної ділянки явищ, створенні розгорненої наукової теорії тощо.

Інтерактивний оптимізаційний метод інформаційного пошуку – рух за бібліографічними посиланнями.

Можливість його використання полягає й у специфіці літератури, що складається з певного взаємозв'язку фактів, творів, документів та видань. Кожен новий літературний твір – це певною мірою аналіз, оцінка й узагальнення соціальної інформації, вже наявної в літературі. Особливою формою віддзеркалення вказаного взаємозв'язку, спадкоємності літературного розвитку і є бібліографічне посилання.

Метод сходження від абстрактного до конкретного – це процес пошуку за допомогою класифікаційних систем.

Застосовуючи цей метод, спочатку розглядають більш загальну ділянку знань, яка у процесі пошуку поступово звужується та конкретизується.

Усі ці методи дають змогу знайти потрібну інформацію в традиційному інформаційному середовищі. Проте у наш час не менше, можливо, і набагато більш важливі методи і методика пошуку інформації у новому інформаційному середовищі – Інтернеті.

Пошук інформації можна здійснити за допомогою декількох методів, що значно різняться як за ефективністю та якістю пошуку, так і за типом потрібної інформації.

Перші два методи інформаційного пошуку – суцільний і вибірковий – найбільш традиційні, вони є основними у

бібліографічній евристиці.

13.1.1. Суцільний метод

Використовуючи суцільний метод, обстежують усі наявні джерела інформації. Хоч і матимемо величезні витрати праці й часу, цей метод найбільш ефективний. У сучасних умовах цей прийом важко реалізувати навіть для вузькоспеціалізованої теми: наприклад, він нереальний для фактографічного пошуку, оскільки потребував би суцільного читання величезної кількості літератури. Тим паче, що не варто витрачати сили і час на вивчення літературних джерел, що зовсім не стосуються теми і завдань пошуку. Тому вважатимемо цей метод ідеальним крайнім випадком, а всі інші – своєрідними його "репліками", побудованими за певними правилами.

При реалізації суцільного методу постає низка проблем. Найпоширеніший приклад – назва книги або статті не відповідає змістові твору. Потрібно зважати на можливості звернення до літератури, її наявність при здійсненні такого пошуку. В більшості випадків бібліографія містить тільки наявні джерела. Тому, обираючи раціональніший вибірковий метод пошуку літератури, ми маємо взяти до уваги таке: по-перше, потрібно дуже обережно підходити до будь-якого обмеження суцільного методу, особливо до можливості вивчення або не вивчення того або іншого видання; по-друге, слід заздалегідь визначати певні типи джерел і допоміжних матеріалів для обов'язкового перегляду, виходячи з конкретного завдання пошуку літератури.

13.1.2. Вибірковий метод

Раціональніший і реальніший спосіб пошуку літератури – вибірковий метод, тобто обмежено суцільний, або епізодичний метод, але поняття "вибірковий" більше відповідає його суті й особливостям. Співвідношення цих двох методів (суцільного і вибіркового) потрібно розглядати діалектично. Традиційно їх розділяють за характером звернення до тих чи інших джерел (жанрово-видовий критерій пошуку): з одного боку, рекомендують брати за основу головне джерело або факт літератури, видатної наукової роботи; з іншого – наявні довідково-бібліографічні посібники; а дехто спирається на певну комбінацію, взаємодоповнення джерел і довідників.

Іншими словами, ці методи відрізняються лише рівнем входження у систему літератури і її охоплення – або від одиничного факту, документа (джерела і довідники), або певної частини документів (скажімо, фонду бібліотеки, окремого виду літератури чи літератури в цілому). В обох випадках усе зводиться до підбору і вивчення певної сукупності літературних джерел. Такий взаємозв'язок цих методів забезпечує оптимальний з погляду раціональності обсяг пошуку літератури.

Отже, питання полягає в тому, як визначити достатній обсяг літератури для кожного окремого пошуку. Вибірковий метод можна розглядати як додатковий до суцільного, на цій основі можна заздалегідь визначити ті джерела, які немає потреби вивчати. Але якщо у процесі пошуку раптом за якимось посиланнями, непрямими вказівками тощо виявиться, що у цій заздалегідь відхиленій частині літератури є матеріали, які цікавлять нас, то саме тоді й застосовують вибірковий метод. Тому його можна вважати запасним і використовувати саме частково, епізодично, але водночас не обмежуючи пошук якимось одиничним фактом, документом. Можна починати пошук літератури з вибіркового методу через суцільне вивчення найпростіших довідників. Так, однією з вихідних позицій пошуку літератури можуть бути енциклопедичні словники. Але справа не тільки у специфіці початкового джерела або факту. У процесі пошуку літератури потрібно використовувати всю можливу різноманітність обмежень, тобто у будь-якому разі комплексно, всесторонньо визначаючи, конкретизуючи початкове завдання пошуку. Важливо відзначити, що навіть для використання таких елементарних методів, як суцільний і вибірковий, потрібна заздалегідь сформована певна творча, розумова робота. Це і стає причиною використання методів раціональнішого характеру – інтуїтивного і типологічного (рецептурного).

13.1.3 Інтуїтивний метод

Процес пошуку літератури – справа складна, така, що вимагає працьовитості, енергії, різносторонньої обізнаності й особливих, якщо не методів, то хоч би навиків. Уміння знайти потрібний матеріал сьогодні є результатом бібліографічного чуття та

досвіду. У сучасних умовах таке спрямування інтуїтивного методу пошуку літератури до мистецтва, емпірики, до рівня "особливих", "вибраних" осіб не можна вважати виправданим. Потрібно зважати, по-перше, на той факт, що роль інтуїції в творчій діяльності сучасної людини надзвичайно зросла; по-друге, сучасна наука постала перед альтернативою: або освоїти логіку інтуїції, зрозуміти логічний механізм інтуїції, або щоб логіку поглинув ірраціоналізм.

Звичайно, не можна абсолютизувати інтуїтивний метод, як і зменшувати його значення або ігнорувати взагалі. Будь-який творчий процес є певною єдністю, з одного боку, планомірних, потрібних і логічних висновків, що послідовно витікали один з одного, з іншого – збігу випадкових знахідок, нез'ясовних "осяянь", логічних стрибків і переходів. З цього приводу Антуан де Сент-Екзюпері писав: "Теоретик вірить в логіку. Йому здається, ніби він зневажає мрію, інтуїцію і поезію. Він не помічає, що вони, ці три феї, просто переодягнулися, щоб спокусити його, як влюбливого хлопчиська. Він не знає, що якраз цим феям зобов'язаний він своїми чудовими знахідками. Вони існують у ньому під ім'ям «робочих гіпотез», «довільних допущень», «аналогій», і чи може теоретик підозрювати, що, слухаючи їх, він зраджує суворій логіці і послухає наспіви муз...".

Інтуїтивний метод щодо завдань інформаційного пошуку ще недостатньо вивчений, а точніше, не завжди цілеспрямовано усвідомлений. Інтуїція плідна тією мірою, у якій її уточнив і переробив розум. Інтуїтивне "осяяння" може зацікавити, якщо існує в думці людини, якщо воно очищене і входить, принаймні, у сукупність обґрунтованих думок. Наукове дослідження, як і в нашому випадку – інформаційний пошук, це не низка "бачень" або думок, що не підлягають аналізу і перевірці. У творців науки дійсно бувають "природні одкровення" або "осяяння", але вони ніколи не з'являться до виявлення, формулювання і вивчення проблеми.

Закономірно, що інтуїтивний метод має велике значення для фактографічного пошуку. Його роль може бути істотною і під час документального пошуку.

У принципі, інтуїтивний метод під різними назвами – "здогадка", "чуття", "випадковість" – досить розповсюджений у

бібліографічній евристиці. Тут відкриваються різноманітні способи і невичерпні можливості залежно від особистих властивостей бібліографа, його досвіду у швидкості міркування і тому подібне. Цим термінам потрібно надавати не містичного змісту, сприймати як встановлення зв'язків, тільки віддаленіших, які лежать не на поверхні, але реально існують, а не вигадані.

До цього виду інтуїтивних методів пошуку належать також різні "випадковості" в процесі підбирання літератури. Досвід показує, що так звана випадковість насправді виявляється дією певних обумовленостей.

Отже, потрібно постійно виробляти в собі прозірливість у найбільш оптимальному виборі способів пошуку інформації, звернення саме до тих літературних джерел, які найбільш імовірно містять шукані факти. Освоєння і раціональне використання різних методів інтуїції з'являються разом із досвідом, віком, є результатом ретельного вивчення літератури. Цей взаємоперехід від інтуїтивних до раціональних методів і навпаки, багато в чому зумовив подальший розвиток бібліографічної евристики у напрямі використання раціональніших методів інформаційного пошуку.

13.1.4. Типологічний, або рецептурний метод

Різноманіття можливих завдань інформаційного пошуку можна узагальнити в певну їх типову сукупність. Їх використання дасть змогу для кожного інформаційного пошуку окреслити певний спосіб і документальні засоби (довідники, джерела) для роботи, тобто створити модель вирішення цього пошукового завдання.

Оптимальною типовою моделлю для розробки різних "рецептів" інформаційного пошуку сьогодні є бібліографічний опис. Згідно з ДСТУ ГОСТ 7.1–2006, він є сукупністю відомостей про будь-який документ або його частину, що дають можливість його ідентифікувати, а також отримати уявлення про його зміст, читацьке призначення, обсяг довідкового апарату і т. д. Як бачимо, основне призначення бібліографічного опису – ідентифікувати видання з-поміж усіх інших. Латинське слово "ідентифікація", яке неточно перекладається українською мовою, якраз і відображає процес виявлення, пізнання шуканого

видання (документа) через уподібнення, ототожнення, порівняння його з іншими на основі певних ознак (характеристик). Останні і складають бібліографічний опис, усі елементи якого (передусім обов'язкові) є поштовхом для встановлення і використання різних заздалегідь фіксованих схем ("рецептів") у процесі інформаційного пошуку.

У будь-якому разі типологічний (рецептурний) метод відображає в собі і привносить у процес інформаційного пошуку певний раціональний початок, планомірність. Водночас він показує важливість відправної, початкової точки цього складного процесу. Її можна обрати і на основі інтуїції, і на основі свідомого, цілеспрямованого роздуму, робочої гіпотези, рецепту.

З огляду на це слід наголосити на взаємозв'язку методів евристики з такими строгими методами логіки, як індуктивний і дедуктивний.

13.1.5. Основні схеми і приблизні "рецепти" інформаційного пошуку

Розглянемо схему рецептурного інформаційного пошуку з використанням різних областей та обов'язкових елементів бібліографічного опису. Методика дещо відрізнятиметься залежно від типу пошуку – документального (посібників, джерел), чи фактографічного (фактів, ланки, якої не вистачає, бібліографічного опису). Нижче вказано поля бібліографічного опису та дії, які потрібно вжити в кожному випадку.

І. Область заголовка і відомостей про відповідальність

Документальний пошук. *На певну тему* – слід звернутися до вже наявних посібників з аналогічною тематикою, видань державної бібліографії, огляду літератури на потрібну або наближену тему. *Певного автора (авторів)* – потрібно опрацювати особисті бібліографії цього автора (універсальні або галузеві) енциклопедії, біобібліографічні словники, іменні покажчики видань державної бібліографії, зібрання творів цього автора. *Певного жанру (виду видання)* – слід звернутися до видань державної бібліографії, де вказана література відповідного жанру (наприклад, при пошуку складеного списку літератури на обрану тему – до двотижневика "Державний бібліографічний покажчик України", при пошуку книг – до "Книжкового

літопису", і т. д.). *Видань однієї з іноземних мов* – варто звернутися до видань книжкових палат країн СНД, покажчика книг і брошур, виданих іноземними мовами в "Державному бібліографічному покажчику України", інформаційних видань закордонних організацій НТІ з відповідною тематикою.

Фактографічний пошук. *Заголовки конкретного джерела* – варто переглянути відповідний том "Літопису книг Державного бібліографічного покажчика України", персональну бібліографію відповідного автора або, знаючи інші бібліографічні відомості про нього (назва, рік видання), шукати потрібну ланку у виданнях державної бібліографії; у разі перекладу – покажчик перекладів з іноземних мов. *Конкретного автора* – слід звернутися до бібліографічного словника, каталога заголовків бібліотеки, наочного покажчика у виданнях державної бібліографії, до інших посібників, використовуючи відомі джерела бібліографічного опису. Біографічні дані про автора можна знайти у біобібліографічних виданнях, довідково-енциклопедичних виданнях, передмовах або післямовах до видань цього автора. *Факту рецензування обраної книги* – така інформація подана у "Літописі рецензій"; але потрібно знати автора, назву книги, прізвище рецензента, назву газети, в якій надрукована ця рецензія.

II. Область видання

Документальний пошук. *Другого і подальших видань твору* – слід звернутися до персональної бібліографії, бібліографічного словника, видань державної бібліографії, зібрань творів його автора.

Фактографічний пошук. *Року і місця першого (другого та подальших) видань твору, відмінності їх за змістом й іншим характеристики* – такі відомості містять зібрання творів цього автора, його персональна бібліографія, бібліографічний словник, посібники державної бібліографії.

III. Область вихідних даних

Документальний пошук. *Виданих у якомусь регіоні (країні, місті, області)* – потрібно звернутися до тематичних планів випуску літератури видавництвами цього регіону, каталогів відповідних видавництв, видань краєзнавчої бібліографії,

бібліографічного словника діячів регіону, географічних покажчиків видань державної бібліографії. *Надрукованої видавництвом інформації (видавничою організацією)* – слід звернутися до тематичних планів випуску, ретроспективних каталогів (репертуарів) відповідних видавництв (друкарень). *Виданою за певний період, у певну епоху* – слід звернутися до поточних або ретроспективних покажчиків літератури, планів випуску і каталогів відповідних видавництв.

Фактографічний пошук. *Місця видання твору* – потрібно шукати, використовуючи вже відомі бібліографічні дані; також у виданнях персональної бібліографії, біобібліографічних словниках, допоміжних покажчиках видань державної бібліографії. *Видавництва (видавничі організації)* – треба працювати, використовуючи вже відомі бібліографічні дані, тематичні плани випуску (алфавітний покажчик імен і заголовків), іменний покажчик видань державної бібліографії. *Роки видання* – можна знайти, використовуючи вже відомі бібліографічні дані, поточні й ретроспективні покажчики державної бібліографії, біобібліографічні словники, персональні бібліографії відповідного автора.

IV. Область кількісної характеристики

Документальний пошук. *З наявністю ілюстрацій (фотографій, креслень, карт тощо)* на тему, яка Вас цікавить, – варто опрацювати видання державної бібліографії; тематичні плани випуску відповідних видавництв, інформаційних видань, довідково-інформаційних фондів, органів НТІ, бібліотек. *З наявністю додатків (альбомів, грамплатівок, мікрофільмів, атласів схем і т. д.)* – потрібно переглянути видання державної бібліографії, тематичні плани випуску відповідних видавництв, інформаційні видання органів НТІ.

Фактографічний пошук. *Наявність портрета певного автора* – можна знайти, використовуючи вже відомі бібліографффічні дані до довідково-енциклопедичних видань, зібрань творів, ювілейних видань цього автора, іконографічних фондів бібліотек й органів НТІ. *Наявність ілюстрованого зображення історичної події, процесу, машини і т. ін.* – слід звернутися, використовуючи вже відомі бібліографічні дані, до ілюстрованих видань творів автора, за Вашою темою, видавництва і

т. ін.; іконографічних фондів бібліотек і органів НТІ.

V. Область серії

Документальний пошук. *Окремого видання, що входить до цієї серії,* – слід звернутися до покажчика серій і серійних видань "Державного бібліографічного покажчика України", літописів періодичних видань, які друкуються й інших джерел. *Усіх видань, які входять до однієї серії,* – потрібно шукати у виданнях державної бібліографії, тематичних планах випусків або ретроспективних каталогах відповідного видавництва. *Серійних видань і серій за певною темою, видом (книжна, періодична), для окремої категорії читачів (масова, спеціальна) тощо* – треба опрацювати тематичні плани випусків відповідних видавництв, покажчики газет зі спеціальної тематики і для певних читацьких груп у "Літописі періодичних видань. Державного бібліографічного покажчика України".

Фактографічний пошук. *Відомостей про серійність видання* – потрібно звернутися до покажчика серій і серійних видань, використовуючи вже відомі бібліографічні дані. *Бібліографічних відомостей про певне серійне видання, серії в цілому* – можливо, входять до покажчика серій і серійних видань, літописів періодичних видань, що продовжуються, планів випуску відповідних видавництв.

VI. Область приміток

Дає змогу уточнити і конкретизувати завдання пошуку літератури, даючи можливість скласти чіткіше уявлення про потрібне Вам видання – при використанні елементів цієї області слід звертатися до видань державної бібліографії, інформаційних видань органів НТІ, де область приміток часто розширена до анотації.

VII. Область міжнародного стандартного книжкового номера (ISBN), палітурки, ціни, тиражу

ISBN, як і Міжнародний стандартний серіальний номер (ISSN), який привласнюють книжковим серіям, періодичним виданням, що продовжуються, дає змогу знаходити, користуючись міжнародним або національним каталогом таких номерів, будь-яке видання, оминувши його бібліографічний опис, оскільки ISBN і ISSN індивідуальні для кожної книги або

серіального видання. Крім того, знаючи кодове значення цифр, що утворюють такий номер, можна на основі ISBN або ISSN визначити країну і видавництво, що надрукувало цю книгу або журнал, рік видання й інші бібліографічні відомості.

13.1.6. Індуктивний метод

Загалом індукція (індуктивний метод) у широкому розумінні слова (від латин. inductio – наведення) – це форма мислення, за допомогою якої думку можна спрямувати до якогось загального правила, положення, яке властиве усім одиничним предметам певного класу. Процес пізнання з використанням індукції здійснюється від одиничних думок і фактів до загальних правил, у яких виражена загальна закономірність.

Щодо нашого завдання, то можна говорити про особливий різновид індуктивного методу. Насправді у процесі інформаційного пошуку за темою нашого посібника ми могли б сформулювати власні думки щодо цього:

1) у нашій країні існують видання, присвячені питанням інформаційного пошуку; має діяти видавництво, що спеціалізується на їх випуску, такі видання зазначені у планах випуску або зведених каталогах цього видавництва (наприклад; видавництво “Книга”);

2) у НТЦ нашої країни існує орган (центр), який надає інформацію про видання на тему, що нас цікавить, така інформація, як правило, випускається у вигляді реферативних журналів. Це видавництво, що друкує РЖ “Інформатика”;

3) У Національній книжковій палаті зареєстровано усі видання, які друкують у нашій країні, інформація про це публікується у відповідних бібліографічних виданнях.

Отже, узагальнені дані з питань інформаційного пошуку ми можемо отримати: у планах випуску і зведених каталогах видавництва “Книга”, РЖ “Інформатика”, у виданнях державної бібліографії (“Державний бібліографічний покажчик України” і т. ін.). Отже, завдяки вивченню цих довідників за певний період ми можемо скласти вагомий список літератури з питань, що нас цікавлять. Іншими словами, від виявлення і пошуку окремих

фактів за допомогою індуктивного методу ми переходимо до складання загального, більш-менш повного списку (певного обсягу джерел) з обраної теми.

У нашому прикладі ми використовували такий вид індуктивного методу, як неповна індукція, отже підсумковий результат в такому разі не охоплює всі можливі одиничні джерела і не заснований на їх вивченні.

Іншим видом індуктивного методу є повна індукція. Загальний висновок про наявність шуканих джерел (фактичних відомостей) з питання, що цікавить, тоді можна робити на підставі вивчення усіх одиничних джерел. Такий максималізм, як правило, дає достовірніші та повніші узагальнення у процесі інформаційного пошуку. Але метод повної індукції, як і суцільний метод, в інформаційному пошуку майже не використовують.

Як бачимо, метод індукції тісно пов'язаний з охарактеризованими вище суцільним (повна індукція) та вибірковим методами (неповна індукція), а також з окремими різновидами інтуїтивного методу (наприклад, швидке ототожнення джерела, прискорений висновок, схильність до синтезу). Унаслідок використання індуктивного методу інформаційного пошуку ми отримуємо і своєрідний "рецепт" вирішення поставленого конкретного завдання. Відмінність і переваги індуктивного методу порівняно зі вказаними полягає в тому, що: по-перше, індуктивний метод більш раціональний, логічно строгий; по-друге, він тісно пов'язаний із фактами, практикою, з дійсною ситуацією в інформаційній діяльності. Не випадково основоположник цього методу в науці нового часу – Ф. Бекон – наголошував: "Якщо ми хочемо проникнути в природу речей, то ми усюди звертаємося до індукції... Бо ми вважаємо, що індукція – це справжня форма доказу, що оберігає відчуття від всякого роду помилок, близько стежить за природою, граничить і майже зливається з практикою".

13.1.7. Дедуктивний метод

Дедукція (від латин. deductio – виведення) в широкому значенні слова – це така форма мислення, коли нову думку виводять винятково за допомогою логіки (тобто за законами

логіки) з попередніх думок. Іншими словами, змістом дедукції як методу пізнання є використання загальних положень при дослідженні конкретних явищ, причому спрямованість думки від загального до конкретного може існувати не тільки в окремому акті пізнання, яким є висновок. Така спрямованість думки може діяти також при дослідженні певної ділянки явищ, створенні розгорненої наукової теорії тощо.

Особливий різновид дедуктивного методу використовують у процесі інформаційного пошуку. Звертаючись знову до вказаного типового прикладу про пошук літератури на тему нашого посібника, ми, застосовуючи дедуктивний метод, можемо зробити висновки. По-перше, процес пошуку інформації та його теорія (бібліографічна евристика) як окремий випадок (тема) входять у зміст книжкової справи й бібліографії та їх теорій (книгознавства, бібліографоведення), а також інформатики. Отже, саме в довідниках (інформаційних виданнях), де відображений зміст цих ділянок суспільної діяльності й науки, ми можемо знайти відомості, що нас цікавлять, на тему пошуку. По-друге, функціонують інформаційні органи в УНТЦ, які займаються інформаційним обслуговуванням цих сфер суспільної діяльності (бібліографія, науково-інформаційна діяльність): УНТЦ, Інформаційний центр з проблем культури і мистецтва. По-третє, ми маємо визначити, які довідники цих організацій загалом можуть відображати літературу, яка нас цікавить. Такими є: РЖ “Інформатика”, бібліографічний покажчик Інформцентру з проблем культури і мистецтва “Нова українська й іноземна література з бібліотекознавства і бібліографії”. Нарешті, по-четверте, через суцільне або вибіркове (використовуючи зміст, допоміжні покажчики і тому подібне) вивчення вказаних довідників ми можемо сформувати потрібний список літератури на тему інформаційного пошуку.

Прикладу, котрий ми навели, досить, щоб підтвердити наявність тісного взаємозв’язку між індуктивним і дедуктивним методами. Водночас деяка абсолютизація індуктивного методу почалася вже у Ф. Бекона, хоча він і застерігав від того, щоб не уподібнюватися “емпірикові-мурашці”, але і не бути схожим на

“павука-раціоналіста”, який із власного розуму може зіткати хитру філософську павутину. На думку Ф. Бекона, філософ має бути схожий на бджолу, яка збирає нектар у полях і лугах, а потім виробляє з неї мед. Характерно, що вже у давнину і книжника порівнювали з бджолою. В усякому разі, процес інформаційного пошуку не зводимо тільки або до індукції – вивчення літератури від одиничних фактів (джерел) до їх об’єднання (зведення), або до дедукції – вивчення літератури від узагальнених фактів (звідних джерел, довідників) до одиничних. Індукція і дедукція найтісніше взаємопов’язані, доповнюють одна одну. Як індуктивне вивчення припускає використання якихось загальних положень, узагальнювальних джерел (довідників), тобто має елементи дедукції, так і навпаки: дедукція не можлива без узагальнень і висновків, які отримують індуктивним методом.

Отже, ефективність інформаційного пошуку великою мірою залежатиме від правильного, раціонального поєднання індуктивного і дедуктивного методів, а також, своєю чергою, їх поєднання з іншими відомими методами бібліографічної евристики.

13.1.8. Пошук за бібліографічними посиланнями

Можливість використання цього методу залежить і від особливостей характеру літератури, до яких входить певний взаємозв’язок фактів, творів, документів і видань. При використанні раціональних й інтуїтивних методів інформаційного пошуку, ми і намагаємося ймовірнісним, гіпотетичним способом виявити цей взаємозв’язок і встановити шукані факти і джерела. Головна особливість методу інформаційного пошуку за бібліографічними посиланнями в тому і полягає, що тут використовують не гіпотетичний, а реальний взаємозв’язок літератури у процесі її історичного розвитку. Кожен новий літературний твір – це певною мірою аналіз, оцінка й узагальнення соціальної інформації, вже наявної в інших джерелах. Специфічною формою віддзеркалення такого взаємозв’язку, спадкоємності літературного розвитку і є бібліографічне посилання (про що йшлося вище, на с. 47).

Для нас важливо, що під час інформаційного пошуку наявність бібліографічних посилань на першому етапі відкриває такі можливості: по-перше, обмежитися одним відомим або знайденим джерелом з питання, що цікавить нас. Використовуючи систему вказаних у ньому бібліографічних посилань, ми зможемо пізніше сформувати відповідний список джерел. Цей список, як ми знаємо, може міститися вже у власне першоджерелі, оформлений у вигляді бібліографічного покажчика списку (огляду). Тепер, послідовно вивчаючи систему бібліографічних посилань у кожному із вказаних у списку джерел, ми ретроспективно визначаємо шукану літературу з нашого питання. Головна вимога – у процесі пошуку, якщо можливо, брати за основу найсвіжіше видання. По-друге, можна обмежитися одним відомим або знайденим списком (покажчиком) шуканої літератури, що публікується в критико-бібліографічних розділах періодичних видань, в інформаційних, довідково-енциклопедичних та інших виданнях, які складають у бібліотеках тощо. Нарешті, по-третє, можна відштовхуватися не від бібліографічного посилання (списку) в цілому, а від окремих його елементів: наприклад, нам відомий автор цієї книги, тоді, користуючись алфавітним каталогом певної універсальної бібліотеки, іменними допоміжними покажчиками до інформаційних видань (так, як видання державної бібліографії), до окремих книг, біобібліографічних словників, енциклопедій і тому подібне, ми встановлюємо інші видання цього ж автора. Тоді за системою бібліографічних посилань цих видань знаходимо праці інших авторів з питання, що цікавить нас.

В умовах сучасного бурхливого зростання потоку інформації значення методу пошуку за бібліографічними посиланнями суттєво зростає. Для його спрощення розроблений і видається спеціальний вид бібліографічної довідкової літератури – покажчик цитованої літератури. У наший країні функції такого покажчика певною мірою виконує двотижневик “Державний бібліографічний покажчик України”, у якому крім інформації про окремо видані бібліографічні огляди (списки), бібліографічні періодичні видання, зазначені також відомості

про найбільші внутрішньокнижні та внутрішньожурнальні бібліографічні покажчики (списки, огляди), прикнижні та пристатейні бібліографічні списки. За допомогою іменного покажчика у щорічниках можна використовувати при пошуку такі огляди, як покажчик цитованої літератури. У будь-якому разі метод пошуку літератури за бібліографічними посиланнями – один із найпрактичніших і доступних для кожної бібліографії.

Та все ж, незважаючи на своєрідність і певні переваги кожного з охарактеризованих вище методів інформаційного пошуку, використання тільки одного з них недостатньо. Оптимальним буде застосування певної їх сукупності. Але для цього потрібно відшукати якусь універсальну методологічну стратегію, своєрідну модель. Саме такою моделлю, в якій діалектично взаємозв'язані всі можливі методи бібліографічної евристики, і є спосіб (метод) сходження від абстрактного до конкретного.

13.1.9. Метод сходження від абстрактного до конкретного

Згідно з визначенням К. Маркса, в найзагальнішому під методом сходження від абстрактного до конкретного (і навпаки) розуміють спосіб, "за допомогою якого мислення засвоює собі конкретне, відтворює його як духовно конкретне". Зауважимо також, що К. Маркс характеризував цей метод як "правильний в науковому відношенні. Більш того, всі інші наукові методи включаються в цей універсальний діалектико-логічний метод на правах часткових випадків".

Остання думка повністю стосується й методів бібліографічної евристики. Досить сказати, що майже всі охарактеризовані вище методи інформаційного пошуку обмежуються лише першим ступенем мислення: абстрактне (обмежена інформація про предмет пошуку) – реальне конкретне (література як реальна сукупність документів, видань тощо). Виняток можуть становити тільки індуктивний і дедуктивний методи інформаційного пошуку. Знову ж таки, лише у своєму взаємозв'язку вони певною мірою наближаються до методу сходження від абстрактного до конкретного, відображаючи ступені багатократного пізнання від одиничного до загального і навпаки. Усі ці методи інформаційного пошуку можна вважати

елементарними – кожен з них односторонній, недостатній для всеосяжного дослідження літератури, тому ефективний пошук досягається лише за використання певної їх сукупності. Але саме знання суті інформаційного пошуку методом сходження від абстрактного до конкретного дає цю визначеність.

На переваги використання цього методу вказує і складність власне літератури як конкретного, реального об'єкта інформаційного пошуку. Одне й те саме джерело, факт має в літературі виняткове різноманіття зв'язків і відношень, які дуже важко простежувати. Для прикладу, достатньо перерахувати ті перетворення, які обов'язково стаються з кожним літературним джерелом (книгою) в сучасній системі інформаційного спілкування, суспільній діяльності в цілому.

1. Ще до видання книга як літературне джерело (твір, документ) може стати предметом: а) доповіді на різних конференціях, –відображаючись у матеріалах цих конференцій; б) звіту про результати науково-дослідної або проектно-конструкторської роботи – буде занесений до довідково-інформаційного фонду певної установи і відображений у довідкових каталогах цього фонду; в) публікації в періодичному виданні (журналі, газеті); г) анотації в тематичному плані випуску якогось видавництва.

2. Після виходу книги у світ інформація про неї міститиметься: ґ) у виданнях державної бібліографії ("Літопис книг: Державний бібліографічний покажчик України"); д) в інформаційних виданнях відповідних органів УНТЦ; е) у разі публікацій у пресі на неї рецензії – у виданні державної бібліографії "Літопис рецензій: Державний бібліографічний покажчик України"; є) у разі її використання і віддзеркалення в прикнижном списку літератури (бібліографічному довіднику) – у щорічнику державної бібліографії "Державний бібліографічний покажчик України"; ж) у звідних каталогах видавництва, що опублікувало її; з) у персональній бібліографії відповідного автора; и) у біобібліографічних словниках письменників й учених.

3. Під час надходження книги у систему книжкової торгівлі,

інформація про неї буде поміщена: і) у різних довідниках книготоргової бібліографії; ї) у рекламних виданнях; й) у каталогах букіністичної книжкової торгівлі.

4. Після надходження книги у бібліотеки інформація про неї буде поміщена: к) у каталогах і різних довідкових картотеках бібліотеки; л) у виданні звідного каталогу бібліотеки.

Як бачимо, тільки обов'язкові (16 – і це ще не всі!) з багатьох можливих варіантів віддзеркалення певного джерела або посібника в літературі настільки різні за своїм змістом, формою подачі, систематизації матеріалу й іншими ознаками, що викликають зацікавленість, мають значення при інформаційному пошукові.

До процесу пошуку належать чотири основні етапи. Перший з них передбачає цілеспрямоване обмеження й подальшу конкретизацію загальної мети пошуку тільки пошуком бібліографічних допоміжних матеріалів, які, можливо, містять інформацію про шукані джерела. Такими матеріалами можуть бути бібліографічні видання, видання інформаційних центрів, довідково-енциклопедичні видання, різні каталоги універсальних бібліотек. Щоб встановити конкретну сукупність шуканих посібників на тему пошуку, спочатку ми маємо визначити їхню максимальну сукупність. Обов'язковими абстракціями при вирішенні завдання пошуку на цьому етапі стають різні бібліографічні факти – прізвище автора, назва видання й інші відомості, подані у бібліографічному описі джерела. Можливими прикладами конкретних вирішень завдань пошуку на цьому етапі можуть слугувати варіанти "рецептурного" методу пошуку (с. 132). У цьому процесі застосовують усю можливу різноманітність методів, щоб у результаті отримати потрібну сукупність бібліографічних допоміжних матеріалів.

Усе це є початковою передумовою для другого етапу сходження від абстрактного до конкретного – пошуку літератури до питання, бібліографічного списку на потрібну тему.

Завдання третього етапу – визначити ті літературні джерела, у яких є інформація, що цікавить нас. Отже, тепер можливе

застосування варіанту документального пошуку. Роль абстракцій, або так званих пошукових образів, тут вже виконують не тільки бібліографічні описи (анотації, реферати) можливих літературних джерел (книг, статей і тому подібне), а й власне ці джерела. Потреба у їх залученні визначається, передусім, через вивчення довідкового апарату: видавничої анотації, змісту, передмови, післямови тощо. У результаті – формується шуканий обсяг у потрібних джерел на тему пошуку (не тільки у вигляді списку, а й самих документів і видань).

Четвертий етап передбачає визначення потрібної сукупності фактографічної інформації на задану тему, тобто тепер проводиться фактографічний пошук через читання і конспектування літературних джерел. Як наслідок цієї роботи формується шуканий обсяг фактографічної інформації, або початковий матеріал для бібліографічного обслуговування.

Отже, тільки унаслідок низки сходжень від абстрактного до конкретного (література – допоміжні матеріали – джерела – літературні факти) можна реалізувати конкретне завдання інформаційного пошуку з використанням цього методу.

13.2. Питання для самоперевірки

1. Назвіть основні види інформаційного пошуку.
2. Що таке стратегія пошуку?
3. Метод пошуку – це...
4. Якими методами можна скористатися під час пошуку інформації?
5. Охарактеризуйте суцільний метод інформаційного пошуку.
6. Вкажіть особливості вибіркового методу інформаційного пошуку.
7. Дайте характеристику інтуїтивному методові інформаційного пошуку.
8. Що таке типологічний метод інформаційного пошуку, які його особливості?
9. Індуктивний метод інформаційного пошуку та його наукові можливості.
10. Охарактеризуйте дедуктивний метод інформаційного пошуку.
11. Основні риси методу інформаційного пошуку за

бібліографічними посиланнями.
12. Визначте особливості методу сходження від абстрактного до конкретного.

14. ПОШУК В ІНТЕРНЕТІ ЗА ДОПОМОГОЮ ПОШУКОВИХ МАШИН

14.1. Оцінка якості пошуку

Якість пошуку, як правило, оцінюють за двома параметрами – точністю та повнотою пошуку.

Точність пошуку – частка релевантних документів у результатах запиту пошукової системи.

Повнота пошуку – частка знайдених релевантних документів у загальній кількості релевантних документів пошукової бази даних.

Крім цього можна визначити і релевантність результатів пошуку.

Релевантність – ступінь відповідності знайденого документа пошуковому запиту.

Проте релевантність – це тільки формальна характеристика, яка аж ніяк не вказує, чи відповідають знайдені документи пошуковим потребам користувача. Тоді слід говорити про пертинентність знайденої інформації.

Пертинентність – ступінь відповідності знайденого документа інформаційним потребам.

Ці параметри дещо абстрактні, оскільки неможливо точно визначити саму релевантність. Як показують дослідження, коли релевантність документів оцінюють два спеціалісти, то між двома множинами відібраних релевантних документів спільними виявляються в середньому тільки 40 % релевантних документів. Точність і повнота пошуку не перевищує 65 %, тому якість пошуку не може перевищувати 65 % унаслідок різного сприйняття інформації окремими людьми.

14.2. Пошук дублікатів документів

Простота створення електронних документів через копіювання або просту компіляцію декількох джерел зумовила проблему пошуку та створення копій документів.

Походження копій документів в Інтернеті найрізноманітніше. Один і той самий документ на одному і тому ж сервері може відрізнятися за технічними параметрами: він може бути поданий у різних кодуваннях і форматах, може містити змінні вставки – рекламу, дату та ін.

Велику кількість документів з Інтернету можна копіювати і редагувати – стрічки новин, юридичні документи, прейскуранти магазинів, відповіді на найпоширеніші питання тощо. Популярні типи змін: коректура, реорганізація, ревізія, реферування, розкриття теми і т. ін. Нарешті, публікації можуть бути скопійовані з порушенням авторських прав і змінені задля ускладнення їх пошуку.

Крім того, індексація пошуковими системами динамічних сторінок зумовлює ще один поширений вид документів, які зовні мало відрізняються: анкети, форуми, сторінки товарів в електронних магазинах тощо.

Повторення трапляються і в науковій літературі. Часто автори повторно друкують одні й ті самі результати в різних журналах. За даними опитування американських учених-біомедиків, рівень повторних публікацій становить 4,7 %, а ступінь прямого плагіату – 1,4 %! Очевидно, що з повними повторами проблем особливих немає, досить зберігати в індексі контрольну суму тексту й ігнорувати решту текстів з такою самою контрольною сумою. Проте цей метод не працює для виявлення хоч би трохи змінених документів.

Одним зі способів вирішення цієї проблеми є метод Уді Манбера (Udi Manber), створений у 1994 р., та Андрія Бродера (Andrei Broder), котрий удосконалив цей алгоритм, що отримав назву “шинглу”. Шингл – контрольна сума частини тексту.

Як же працює алгоритм? Для кожних послідовних десяти слів тексту розраховують контрольну суму (шингл). Для цього беруть усі

можливі десятислів'я – їхня кількість у тексті буде на 9 позицій меншою від загальної кількості слів у тексті; потім із цієї сукупності шинглів відбирають лише ті, які діляться, припустимо, на 25. Оскільки значення контрольних сум розподілені рівномірно, критерій вибірки ніяк не залежить від особливостей тексту. Зрозуміло, що повтор навіть одного десятислів'я – вагома ознака дублювання, якщо ж їх багато, наприклад, більше половини, то з певною ймовірністю можна стверджувати, що знайдений документ – копія. Адже збіг одного шинглу у вибірці дорівнює приблизно 25 десятислів'ям у всьому обсязі тексту.

Очевидно, що так можна визначати відсоток копіювання текстів, виявляти всі його джерела тощо. Існують програми, що використовують цей алгоритм для пошуку складників тексту і дублікатів в Інтернеті. Тепер допитливий викладач за бажанням може з'ясувати, звідки списано певний реферат; так можна легко оцінити частку плагіату в будь-якій статті. Варто зазначити, що у деяких країнах ставлення до плагіату дуже серйозне – за це можна поплатитися не тільки репутацією, а й можуть звільнити з університету.

Звичайно, подібні алгоритми існують і в пошукових машинах, котрі відсіюють дублікати або групують подібні сторінки в кластери задля зменшення списку знайдених документів. Як правило, перші десять текстів зі списку видачі тоді відзначається більшою різноманітністю.

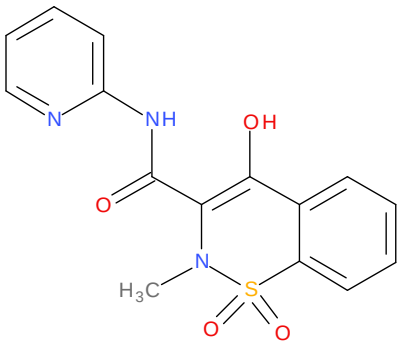
14.3. Проблема вибору пошукової машини

Як уже зазначалося, у світі існує багато пошукових машин, і тут постає питання вибору оптимального варіанту. Перед тим, як вдаватися до конкретних методів пошуку інформації, давайте розглянемо принципи роботи різних пошукових систем. Для прикладу візьмемо нескладну хімічну сполуку – піроксикам, котрий використовують у медицині як протизапальний засіб. Змоделюємо різні форми подачі запитів для цієї речовини та приблизно оцінимо пертинентність одержаних результатів. Припустимо, що нас цікавить загальна хімічна інформація про цю сполуку. Будемо оцінювати відповідність до наших

інформаційних потреб перші 10 посилань у списку результатів пошуку та кількість релевантних результатів, виданих у системі. Як запит використаємо назву піроксикаму українською, російською та англійською мовами, брутто-формулу, хімічні ідентифікатори: SMILE, InChi, InChiKey.

Приклад 32. *Різні форми відображення інформації про хімічну сполуку піроксикам:*

Структурна формула:



Систематична назва – 4-гідроксі-2-метил-N-(піридин-2-іл)-2H-1,2-бензотіазин-3-карбоксаміду 1,1-діоксид;
назва українською – піроксикам;
назва російською – пироксикам;
назва англійською– piroxicam;
Брутто формула – $C_{15}H_{13}N_3O_4S$;
SMILES – OC=2c1ccccc1S(=O)(=O)N(C)C=2C(=O)Nc3ccccn3;
InChI – InChI=1S/C15H13N3O4S/c1-18-13(15(20)17-12-8-4-5-9-16-12)14(19)10-6-2-3-7-11(10)23(18,21)22/h2-9,19H,1H3,(H,16,17,20);
InChIKey – QYSPLQLAKJAUJT-UHFFFAOYSA-N;
CAS Registry Number – 36322-90-4.

Таблиця 3

Залежність кількості релевантних та пертинетних документів від форми пошукового запиту

Пошукова машина	Піроксикам (укр.)	Пироксикам (рос.)	Piroxicam (англ.)	$C_{15}H_{13}N_3O_4S$	SMILES	InChI	InChIKey	CAS #
www.google.com*	94200 (1)	1090000 (1)	1070000 (1)	70320 (7)	3 (3)	2 (2)	5 (4)	231000 (2)
www.bing.com	81 (0)	4380 (1)	404000 (1)	293 (4)	–	1 (1)	–	1560 (2)
www.altavista.com	281 (0)	56200 (0)	1750000 (1)	1700 (0)	–	2 (2)	2 (2)	104000 (2)
www.yahoo.com	530 (0)	56100 (0)	1720013 (0)	1550 (2)	–	2 (2)	2 (2)	104000 (3)
www.yandex.ru	1142 (0)	51000 (1)	99 тис. (1)	6 млн (2)	–	–		624 (2)
www.rambler.ru***	140/1534 (1)	3000/37000	3000 / 75000	6/ 31	160 / 333	–	–	341 / 2 тис.

Пошукова машина	Піроксикам (укр.)	Пироксикам (рос.)	Piroxicam (англ.)	$C_{15}H_{13}N_3O_4S$	SMILES	InChI	InChIKey	CAS #
		(0)	(0)	(1)	(0)			(2)
www.nigma.ru**	4300 (1)	8,3 млн (1)	563 тис (0)	225 (4)	6,7 тис (0)	13 (1)	5 (0)	2200 (2)
www.meta.com.ua***	102/1100 (0)	282/2900 (0)	256/5700 (0)	–	–	–	–	633/1200 (0)
www.scirus.com	7 (0)	40 (1)	39735 (2)	38 (10)	–	–	–	135 (2)

* Проаналізовано попередні запити. Кількість і якість результатів залежатимуть від попередніх запитів.

** Одержано неочікувані результати, які можуть мати користь.

*** Кількість сайтів/кількість документів.

Буквально з перших сторінок зрозумілі ділянки застосування речовини. Цього цілком достатньо, щоб написати вступ.

Із одержаних результатів можна зробити такі висновки:

Кількість знайдених документів різними мовами зростає у рядкові “українська” – “російська” – “англійська”. Це означає, що найбільш повний пошук можна провести тільки англійською мовою. Документів українською мовою досить мало.

При використанні пошукових систем загального призначення, вживаючи простий однослівний запит, на першій сторінці результатів пошуку пертинентних документів практично немає.

Використання брутто-формули хімічної сполуки як пошукового терміна значно підвищує пертинентність результатів.

Більшості пошукових машин не доступна специфіка SMILES – вони не “бачать” у цьому рядку запис хімічної формули.

Використання хімічних ідентифікаторів InChI та InChIKey дає невелику кількість здебільшого пертинентних документів, причому ці ідентифікатори мають чітко вказуватися в документі. Це означає, що пошукові системи не обробляють та не перетворюють назви та формули хімічних сполук.

Використання CAS # дає більш релевантні результати, ніж у разі вказування назв речовин, проте одержана інформація, в основному, має комерційний характер.

Отже, при пошуку інформації про певні хімічні сполуки бажано попередньо визначити хоча би брутто-формулу та хімічні ідентифікатори InChI та InChIKey, а потім уже пошукати за іншими ключовими словами.

14.4. Питання для самоперевірки

1. Що таке точність пошуку?
2. Повнота пошуку – це...
3. Дайте визначення релевантності.
4. Що таке пертинентність?
5. Чим релевантність відрізняється від пертинентності?
6. Причини виникнення дублікатів в Інтернеті.
7. Що таке “шингл”?
8. Опишіть механізм пошуку дублікатів за допомогою “шинглів”.
9. Чи залежить кількість знайдених документів від мови поданого запиту?

10. Розташуйте мови запитів у порядку зростання кількості знайдених документів.
11. Які прийоми сприяють підвищенню релевантності результатів пошуку?

15. РЕАЛІЗАЦІЯ РІЗНИХ МЕТОДІВ ПОШУКУ ІНФОРМАЦІЇ В ІНТЕРНЕТІ

Інтернет поступово став частиною нашого повсякденного життя. Кількість сайтів постійно зростає, і все менше стає людей, які при слові “мережа” здивовано знизують плечима. Інтернет – це сучасний, зручний засіб комунікації та обширне інформаційне джерело. У більшості випадків користувачеві доводиться використовувати спеціальні пошукові інструменти, найчастіше – це пошукові системи і каталоги.

Однією з основних проблем повнотекстового інформаційного пошуку є проблема неоднозначного вибору термінів, вживаних користувачами пошукових систем у пошукових запитах. Ця проблема полягає в тому, що користувачі пошукових систем часто застосовують для опису ключових понять терміни, відмінні від тих, які використовували автори для опису тих самих понять у текстах сторінок. Статистика показує, що в будь-якому разі дві різні людини використовують один і той самий термін для опису одного поняття менше, ніж у 20 % випадків. Ця проблема постає ще гостріше для коротких пошукових запитів тому, що перевагою довших за обсягом запитів є більша ймовірність того, що важливі терміни з шуканих документів потраплять у пошуковий запит. Проте користувачі зазвичай не схильні формулювати довгі пошукові запити або використовувати спеціальні засоби, що дають змогу точніше його формулювати (наприклад, мови пошукових запитів).

З огляду на цей фактор ключовими завданнями в теорії інформаційного пошуку стали:

– розширення пошукових запитів, що полягає в додаванні до початкового запиту синонімів або словоформ ключових термінів запиту без зміни значень початкового запиту;

– уточнення (або переформулювання) пошукових запитів, що полягає у зміні початкового запиту за допомогою перегляду ключових слів з релевантних документів задля уточнення

значеннєвого змісту запиту і, як наслідок, точності пошуку.

Можна зробити висновок, що більшість користувачів Інтернету застосовують індуктивний метод пошуку інформації, або метод спроб та невдач.

Наприклад, у пошуковій системі Рамблер, щоб знайти сайт, де вказані хімічні властивості диклофенаку, можна починати із запиту “диклофенак”. Якщо потрібний сайт не зазначений у списку знайдених документів (принаймні, на першій сторінці видачі), запит треба розширити і переформульовувати на “диклофенак властивості”. І так доти, поки у списку знайденого не з’явиться потрібний сайт.

Який семантичний механізм “розширення” й “уточнення” запиту? Багато користувачів шукають різні парадигматичні зв’язки між словами для уточнення запитів, зокрема синоніми і варіанти (наприклад, заміна запиту “диклофенак” на “вольтарен”). Часто до запиту додають слово (або декілька слів), що позначають більш загальне, родове поняття. Наприклад, така ситуація виникає при розширенні запиту “акваріум” (може стосуватися акваріума для рибок, і назви музичного колективу “Акваріум”) за використання більш загального поняття “музика” – акваріум музика.

На жаль, досить важко самостійно переформулювати запит так, щоб це привело до підвищення релевантності знайдених документів. Далі ми розглянемо різні підходи до пошуку інформації, які дають змогу з мінімальними тимчасовими втратами уточнювати запити й отримувати потрібну інформацію. Треба сказати, що користувач Інтернету рідко обирає якийсь із описаних вище методів пошуку інформації без допоміжних елементів. Найчастіше це модифікації індуктивного методу, котрі полягають у підбиранні ключової фрази.

15.1. Дедуктивний метод

Під час пошуку інформації з використанням дедуктивного методу будемо ґрунтуватися на таких тезах:

– згідно із законом Ципфа, слова, що характеризують сенс тексту, розміщені в середній частині частотного списку тексту;

– пошукові системи також керуються законом Ципфа й присвоюють максимальний рейтинг словам із середини частотного списку тексту.

Тому, якщо ми візьмемо певний текст (літературний огляд,

статтю, реферат тощо), що відповідає темі нашого дослідження, зробимо його частотний аналіз, виберемо ключові слова для подальшого пошуку із середини частотного списку, то при використанні цих ключових слів у запитах маємо отримати найбільш релевантні документи, які будуть найближчими за темою до вихідного тексту.

Для використання цього методу нам потрібен текст-зразок на задану тему. Для прикладу візьмемо бета-версію (чорновик) цього посібника. Для аналізу використаємо засоби для підрахунку частоти слів*. Після аналізу маємо:

текст містить 11782 унікальні слова, найбільша частота вживання службових слів: “в” – 1345, “і” – 1253, “на” – 707, “що” – 688, “з” – 653 і т. д. Зі списку видалимо службові слова – займенники, частки, прийменники. Тепер розрахуємо ранг частоти вживання слів (слова з однаковою частотою мають однаковий ранг). Побудуємо графік ранг-частота:

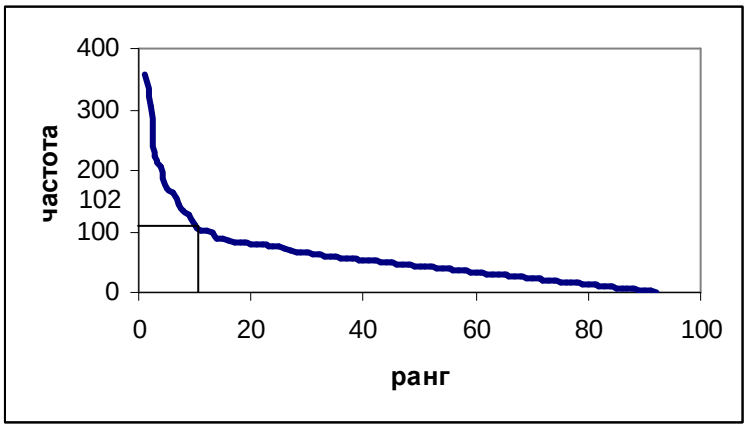


Рис. 16. *Залежність “ранг–частота” для тексту*

Середній діапазон графіка (точка перегину) відповідає

* Потрібні інструменти доступні на веб-сторінці курсу “Наукова інформація з хімії, її пошук та обробка” – http://194.44.187.135/curse/scichem/index.htm

частоті приблизо 102. Згідно із законом Ципфа слова в цьому діапазоні визначають зміст тексту. Виберемо близько 20 слів із цього діапазону нашого списку (звертаємо увагу, що у багатьох пошукових систем довжина запиту обмежена, наприклад у Yandex загальна довжина запиту не може перевищувати 40 слів (разом зі службовими)):

Приклад 33. *Вибрані слова з частотного списку:*

бази, інформація, видань, пошукових, також, система, веб, науки, використання, слова, метод, року, запиту, різних, статей, наукових, issn, літератури.

Сконструюємо з них запит, використовуючи логічне "АБО" для зв'язку слів. Це забезпечить нам найбільш широке охоплення пошукової бази даних:

Приклад 34. *Пошуковий запит на основі вибраних слів:*

бази OR інформація OR видань OR пошукових OR також OR система OR веб OR науки OR використання OR слова OR метод OR року OR запиту OR різних OR статей OR наукових OR issn OR літератури.

Використаємо цей запит у різних пошукових системах загального користування. Будемо аналізувати пертинентність перших десяти одержаних результатів, наскільки вони відповідають темі нашого посібника. Отже:

www.google.com – 434 млн сторінок – результати не пертинентні;

www.yahoo.com – 342 млн сторінок – результати слабо пертинентні – система використала єдине англомовне слово ISSN, тому є результати на цю тему;

www.bing.com – 29 млн сторінок – результати слабо пертинентні, як і в www.yahoo.com.;

www.rambler.ru – 545 млн сторінок – результати не пертинентні;

www.yandex.ru – 188 тис. сторінок – результати пертинентні, зміст відповідає тематиці посібника!

Як ми бачимо, цей метод не завжди дає задовільні результати. Кінцевий результат значною мірою залежить від того, як певна пошукова машина розраховує рейтинг сторінок.

Водночас, в разі успіху потрібні документи можна знайти майже відразу!

15.2. Індуктивний метод

Звичайний метод пошуку, який застосовують користувачі Інтернету – це підбір ключових слів методом спроб та помилок. Саме цей підхід і використовує індукцію, щоб знайти потрібне.

Розглянемо приклад. Нехай нам треба знайти методи аналітичного визначення диклофенаку. Ми вже визначили, що вибрати інформацію з цього питання українською мовою малоймовірно, тому використаємо російську та англійську. Пошукову систему виберемо Google, оскільки вона має найширший обсяг інформації.

Запит 1: “диклофенак определение”.

На першій же сторінці згадується про потенціометричний метод визначення.

Запит 2: “диклофенак потенциометрия”.

Майже всі посилання – на роботи щодо потенціометричного визначення диклофенаку. Спробуємо знайти інші методи.

Запит 3. “методы определения диклофенака”.

Тут ми дізнаємося, що існують екстракційно-фотометричний і хроматографічний методи (на другій сторінці).

Залежно від завдання ми можемо далі підбирати ключові фрази. Успішність пошуку значно залежить від вдалого підбору слів запиту.

15.3. Питання для самоперевірки

1. Наскільки часто різні люди вживають однакові терміни для опису інформації?
2. Які варіанти найчастіше використовують, формулюючи пошукові запити?
3. Які ключові завдання інформаційного пошуку?
4. Які передумови використання дедуктивного методу пошуку інформації?
5. Опишіть процедуру пошуку ключових слів на основі частотного аналізу тексту-зразка.

6. Як потрібно формулювати запит для досягнення максимальної релевантності?
7. Чому різні пошукові системи дають неоднакові результати?

16. ІНДЕКСАТОРИ І КЛАСИФІКАТОРИ

Щоб орієнтуватися у великих обсягах інформації, розроблено різноманітні системи індексів і класифікацій. Їх використання дає змогу так розмістити джерела інформації (фізично на стелажах бібліотек або віртуально на електронних носіях), щоб забезпечити швидкий доступ до потрібно джерела, або швидко ототожнити з об'єктом. Системи індексацій і класифікацій можуть бути і загальні, і спеціалізовані. Класифікація використовується для віднесення джерела інформації до певного класу – за темою, комерційним призначенням; індексація – для ідентифікації (ототожнення) з об'єктом, або для швидкого пошуку. Чітку межу для їх розрізнення часто важко провести, тому обидві категорії ми будемо розглядати разом.

16.1. УДК

Універсальна десяткова класифікація (УДК) – система класифікації інформації, що широко використовується у всьому світі для систематизації творів науки, літератури і мистецтва, періодичних видань, різних видів документів й організації картотек. Усі права на УДК належать організації OCLC (англ. Online Computer Library Center, Inc., – Мережевий комп'ютерний бібліотечний центр), яка удосконалює її відповідно до розвитку науки і техніки. Це велика некомерційна організація, комп'ютери якої об'єднують більше 21 тис. бібліотек у 63 країнах світу. OCLC також займається дослідженнями і технічними розробками в галузі електронних бібліотек.

Уперше систему десяткової класифікації запропонував американський бібліограф Мелвіл Дьюі (Melvil Dewey) 1873 р., опублікована вона 1876 р. На початку XX ст. концепцію УДК розвинули бельгійські бібліографи Полемо Отле (Otlet) та Анрі Лафонтен (Lafontaine).

У нашій країні УДК тривалий час вважали ідеологічно неприйнятною через її "буржуазне" походження і прийняли

тільки у 1962 р. як єдину обов'язкову класифікацію. З 1963 р. у всіх науково-технічних видавництвах, редакціях науково-технічних журналів, в органах науково-технічної інформації, науково-технічних бібліотеках уведено обов'язкове індексування усіх публікацій за Універсальною десятковою класифікацією, тобто всі інформаційні матеріали в галузі природничих і технічних наук публікують з індексами Універсальної десяткової класифікації (УДК).

Перше видання повних таблиць УДК опубліковано французькою мовою 1905 р. Структура УДК з часом відхилилася від початкової схеми М. Дьюї, але у багатьох розділах індекси класів цих систем майже не змінилися. На сьогодні маємо 22 видання таблиць УДК. Ця класифікація одна з найбільш використовуваних у світі: бібліотеки більше 135 країн використовують УДК для організації фондів, обов'язкове присвоєння індексу УДК прийнято в понад 60 країнах світу, таблиці УДК перекладені на більш ніж 30 мов світу.

За цією самою системою організовані довідково-інформаційні фонди (ДІФ). УДК можна використати для організації і вузькоспеціалізованих ДІФ, і багатогалузевих; це єдина міжнародна універсальна система, що дає змогу достатньо детально розкрити зміст довідково-інформаційних фондів і забезпечити швидкий пошук інформації. Можливість єдиної систематизації інформаційних матеріалів робить систему УДК найзручнішою в процесі обміну інформацією між країнами.

УДК відповідає найбільш істотним вимогам, що пред'являються до класифікації: міжнародності, універсальності, мнемонічності, можливості віддзеркалення нових досягнень науки і техніки без якихось серйозних змін у її структурі. УДК охоплює всі галузі знань, її розділи органічно пов'язані так, що зміна одного з них спричиняє зміну іншого.

Десятковою УДК називається тому, що для її побудови використаний десятковий принцип.

Для позначення класів (розділів) застосовано арабські цифри, які абсолютно однозначні для всіх людей незалежно від того, якою мовою вони говорять і яким алфавітом користуються. Мова цифр усім зрозуміла, легко запам'ятовується, тому це робить УДК загальнодоступною

міжнародною системою.

Десятковий принцип структури УДК дає змогу практично необмежено розширювати її завдяки додаванню нових цифр до уже наявних, не руйнуючи системи в цілому. Індекси УДК можуть пов'язуватися один з одним у найрізноманітніших комбінаціях, завдяки чому в схемі класифікації можна відобразити незліченну кількість понять.

Структура УДК ієрархічна. Основу складають десять різних галузей науки та культури. Кожен розділ, своєю чергою, ділиться ще на десять підрозділів; кожен підрозділ ще на десять і т. д. У міру руху ієрархічним деревом, зміст поняття звужується від більш загального до конкретного, так уточнюючи зміст документа.

Кожен клас (перший ступінь ділення) містить групу більш-менш споріднених наук, наприклад, клас 5 – математику і природничі науки, клас 6 – прикладні науки: техніку, сільське господарство, медицину.

Усього існує 10 основних розділів УДК.

16.1.1. Основні розділи УДК:

0. Загальний відділ;
1. Філософські науки. Філософія;
2. Релігія. Атеїзм;
3. Суспільні науки;
4. Вільний з 1961 р., зараз не використовується;
5. Математика. Природничі науки;
6. Прикладні науки. Медицина. Техніка;
7. Мистецтво. Декоративно-прикладне мистецтво. Фотографія. Музика. Ігри. Спорт;
8. Мовознавство. Філологія. Художня література. Літературознавство;
9. Географія. Біографії. Історія;

Хімія подана в більшості у п'ятому й шостому основних розділах УДК.

16.1.2. Підрозділи 5 і 6 основного ділення:

УДК 500–599 Природничі науки і математика;

50. Загальні питання математичних і природничих наук;
51. Математика;

52. Астрономія. Геодезія;
53. Фізика;
54. Хімія. Кристалографія і Мінералогія;
55. Науки про Землю;
56. Палеонтологія. Палеозоологія;
57. Біологія;
58. Ботаніка;
59. Зоологія;
УДК 600–699 Прикладні науки. Медицина. Техніка;
60. Прикладнихі науки. Загальні питання;
61. Медичні науки. Медицина. Пожежна справа;
62. Інженерна справа. Техніка в цілому;
63. Сільське господарство. Лісове господарство. Полювання. Рибне господарство;
64. Домашнє господарство і домоведення. Комунально-побутове господарство;
65. Управління підприємствами. Організація виробництва, торгівлі і транспорту;
66. Хімічна промисловість. Хімічна технологія;
67. Різні галузі промисловості й ремесла. Механічна технологія;
68. Виробництво спеціальних виробів. Точна механіка;
69. Будівництво. Будівельні матеріали. Будівельно-монтажні роботи.

16.1.3. ВЛАСТИВОСТІ І ПРИНЦИПИ УДК

Універсальна десяткова класифікація побудована за систематичним принципом.

В УДК один і той самий предмет зазначений у різних розділах залежно від галузі знань та від аспекту, у якому він розглядається. Наприклад, поняття “нафта” трапляється в декількох розділах. У розділі 547 “Органічна хімія” ми знаходимо поняття “Хімія нафти”; у розділі 553 “Родовища корисних копалин” – “Родовища нафти”; у розділі 622 “Гірська справа” – “Видобування нафти”; у розділі 662 „Вибуховиві речовини. Палива” – “Нафта як паливо” та ін.

Багатоаспектне індексування змісту документів і запитів

забезпечується, окрім власне структури УДК, використанням загальних і спеціальних визначників, що відображають час, місце, мову, параметри, устаткування, процеси і т. д., а також застосуванням певних правил комбінування індексів.

Індекси УДК побудовані так, що кожна подальша цифра не міняє значення попередніх, а лише уточнює, позначаючи більш конкретне поняття.

Наприклад, індекс поняття “Хімічна корозія” 620.193.4 складається так:

6. Прикладні науки;

62. Інженерна справа. Техніка в цілому;

620. Випробування матеріалів. Товарознавство;

620.1. Випробування матеріалів. Дефекти матеріалів. Захист матеріалів;

620.19. Дефекти матеріалів і їх виявлення. Корозія;

620.193. Корозія. Корозійна стійкість;

620.193.4 Хімічна корозія. Дія різних агресивних середовищ.

ТИПОВІ ЗАКІНЧЕННЯ. На початку деяких розділів УДК може міститися перелік цифрових позначень, які починаються багатокрапкою. Це закінчення, які можна за потреби приєднувати до індексів цього розділу для отримання подальшої однотипної стандартної деталізації. Наприклад, при розділі 661.8. “Сполуки металів. Солі. Мінеральні фарби” – на додаток до індексів основного ряду, що позначають сполуки окремих металів:

661.832. Сполуки калію;

661.833. Сполуки натрію;

661.847.9. Сполуки цинку;

– поміщений перелік типових закінчень:

661.8...1. Особливо важливі сполуки;

...2. Оксиди;

...3. Галогенові сполуки. Ціаністі сполуки. Гідриди;

...32. Сполуки хлору;

...321. Хлориди.

Далі зазначено типові закінчення для позначення різних сполук металів з азотом, фосфором, сіркою й іншими

елементами, а також органічних солей.

Будь-яке із закінчень, поданих у переліку, можна використати для отримання індексу сполуки конкретного металу у спосіб безпосереднього приєднання (з дотриманням правила простановки крапки через кожні три знаки) до індексу, що позначає сполуки цього металу. Наприклад, маючи:

з одного боку з іншого боку
661.8...321 Хлориди 661.832 Сполуки калію
532 Сульфат 661.833 Сполуки натрію
і так далі 661.847.9 Сполуки цинку
і так далі
можна утворити індекси:
661.832.321 Хлорид калію;
661.833.321 Хлорид натрію;
661.847.932.1 Хлорид цинку;
661.832.532 Сульфат калію;
661.833.532 Сульфат натрію;
661.847.953.2 Сульфат цинку.

Використовуючи перелік типових закінчень, можна побудувати потрібну деталізацію для сполуки кожного з металів, причому ця деталізація буде стандартною і мнемонічною.

16.1.4. Складові частини УДК

Таблиці Універсальної десяткової класифікації діляться на основні та допоміжні. Крім того, до видань УДК, як правило, входить і третя невід'ємна частина – алфавітно-предметний покажчик.

Ділення таблиць на основні та допоміжні ґрунтується на особливостях відображених у них понять. Як правило, до основної таблиці входять поняття, специфічні для певних галузей науки, техніки, мистецтва і т. д., які наділені тільки властивими для них особливостями.

У допоміжні таблиці (загальних і спеціальних покажчиків) входять поняття, що повторюються, загальні для всіх чи багатьох розділів, або вживані усередині одного розділу. Ці поняття, що приєднуються до основних, уточнюють їх зміст або форму. При цьому загальні визначники містять поняття, вживані

у всіх або багатьох розділах, а спеціальні – використовувані лише в межах одного чи декількох розділів, близьких за змістом.

Класи 5 і 6 тісно взаємозв'язані, і під час індексування часто виникають труднощі при виборі індексу. У цих випадках слід керуватися тим, що клас 5 "Математика. Природничі науки" відображає питання теоретичного змісту, дослідження загальних законів фізики, хімії, біології і т. д., а клас 6 "Прикладні науки. Медицина. Техніка" присвячений питанням практичного використання цих законів, втіленню їх у техніці, медицині, сільському господарстві.

З погляду вченого, УДК не має суттєвого значення. Доведено на практиці, що близько 40 % видань мають неправильний індекс. Присвоєння правильного індексу УДК може мати визначальне значення у подальшій долі видання. У бібліотеці книги та журнали виставляють згідно з їх класифікаційними індексами. У разі присвоєння неправильного індексу книга потрапить не на відповідний стелаж, і може залишитися невідомою для читачів.

Для правильного присвоєння індексу УДК потрібно звертатися до спеціалістів-бібліографів. У Волинському національному університеті імені Лесі Українки УДК присвоюється в університетській бібліотеці. Ця послуга платна.

16.2. ББК

Бібліотечно-бібліографічна класифікація (ББК) – робочі таблиці, призначені для організації бібліотечних фондів, систематичних каталогів і картотек. Перший варіант таблиць розроблений та опублікований у СРСР в 1960–1968 рр. у тридцяти томах під назвою "Бібліотечно-бібліографічна класифікація. Таблиці для наукових бібліотек". У їх розробці брали участь близько 800 учених. Згодом з'явилися також спеціальні видання таблиць для масових бібліотек, обласних, дитячих і шкільних бібліотек.

Класифікація ББК тематична, має систематичний характер, і багато в чому нагадує класифікацію УДК.

Таблиці ББК призначені для організації бібліотечних фондів, систематичних каталогів і картотек.

16.2.1. Основи поділу ББК

В основу розділення таблиць на відділи і розділи покладено

різноманітні критерії: об'єкт вивчення, метод дослідження, мета пізнання, структура об'єкта, його властивості, процеси, відношення, територія, історичний період і т. д. На подальших ступенях ділення іноді застосовується розташування понять за абеткою найменувань (алфавіт країн у межах частини світу й ін.). Водночас у ББК передбачена можливість характеризувати одне і те саме поняття на основі різних ознак, що дає можливість більш глибоко, розкрити зміст опублікованих творів.

16.2.2. Структура ББК

Таблиці ББК для масових бібліотек складаються з декількох видів: основних, загальних, територіальних і спеціальних типових поділів. Усі разом вони складають розгорнуті таблиці.

16.2.3. Основні розділи таблиць ББК

Основний ряд (перші поділи) ББК очолює відділ "Загальнонаукові й міждисциплінарні знання". Наступні відділи охоплюють три основні галузі наукових знань: природничі науки, прикладні науки (техніка, сільське господарство, медицина), суспільні і гуманітарні науки. Замикає основний ряд ББК відділ "Література універсального змісту". Другі, треті, четверті і так далі рівні класифікації утворені через розчленовування поділів першого рівня (основного ряду) на підлеглі групи наук, галузі діяльності, окремі науки, проблеми, теми. У таблицях ББК для масових бібліотек окремі природничі та суспільні науки, які в таблицях для наукових бібліотек подані в основному ряді, перенесені на другий ступінь поділу і підпорядковані узагальнювальним класам.

Таблиці ББК для масових бібліотек відрізняються від таблиць для наукових бібліотек. Перший і другий ряд поділу таблиць ББК для масових бібліотек в цілому відповідає основному (першому) ряду таблиць для наукових бібліотек. Індекси ББК для масових бібліотек цифрові, а для наукових першим знаком є велика буква російського алфавіту.

Таблиця 4

Основні розділи ББК

Масові бібліотеки	Наукові бібліотеки	Розділ ББК

1	2	3
0		Основні відомості про ББК (Бібліотечно-бібліографічну класифікацію)

Закінчення таблиці 4

1	2	3
1	А	Загальнонаукові та міждисциплінарні
2		Природничі науки
20	Б	Природничі науки в цілому
22	В	Фізико-математичні науки
24	Г	Хімічні науки
26	Д	Науки про Землю (геодезичні, геофізичні, геологічні й географічні науки)
28	Е	Біологічні науки
3	Ж/О	Техніка. Технічні науки
4	П	Сільське і лісове господарство. Сільськогосподарські й лісогосподарські науки
5	Р	Охорона здоров'я. Медичні науки
6/8		Суспільні та гуманітарні науки
60	З	Суспільні науки в цілому
63	Т	Історія. Історичні науки
65	У	Економіка. Економічні науки
66	Ф	Політика. Політичні науки
67	Х	Право. Юридичні науки
68	Ц	Військова справа. Військові науки
70/79	Ч	Культура. Наука. Освіта
80/84	Ш	Філологічні науки. Художня література
85	Щ	Мистецтво
86	Э	Релігія. Містика. Вільнодумство
87	Ю	Філософія
88	Ю	Психологія
9	Я	Література універсального змісту

Міждисциплінарні науки в ББК умовно відносять до однієї з них (агробіологія – до сільського господарства, біогеохімія – до біології), а від іншої (або інших) дається посилання. Комплексну науку поділяють на частини, кожна з яких належить до

відповідної “материнської” науки (мікробіологія – до біології, медична мікробіологія – до медицини тощо), а від загальної науки даються посилання на її галузі та конкретні дисципліни.

База індексації ББК – змішана (цифрова і буквена). Окрім арабських цифр використовують: крапку, двокрапку, дефіс, круглі дужки, скісну риску. У таблицях загальних і територіальних типових поділів застосовують рядкові та прописні букви російського алфавіту. Поєднання усіх знаків бази індексації складене у таку систему, в якій кожному знакові належить своє місце, призначення і смислове навантаження.

16.2.3.1. Послідовність індексів

У ББК закріплена певна умовна послідовність рубрик у таблицях і роздільників у каталозі. Порядок розташування індексів на одному ступені поділу такий:

– індекси загальних типових поділів;
– індекси територіальних типових поділів;
– індекси спеціальних типових поділів;
– індекси з двокрапкою;
– індекси основної таблиці з буквеними позначеннями;
– індекси основної таблиці з цифровими позначеннями;
– алфавітно-предметний покажчик.

Якщо одне і те саме поняття відображене в покажчику синонімами, прямими й інверсованими формулюваннями, для гніздових рубрик вибирають найрозповсюдженіший термін, а від них до цього терміна даються посилання. Посилання до гніздових рубрик подані без вказування індексів, оскільки гніздові рубрики містять не один, а декілька індексів, що належать до рубрики.

Структура відділу 24 “Хімічні науки” дещо відрізняється від повного та обласного варіанту ББК. Але вона логічна, і тому ухвалено рішення її зберегти, давши розгорнені методичні вказівки для багатьох поділів. Особливо важливі доповнення, внесено до розділу 24.5 “Фізична хімія. Хімічна фізика”, де відображено такі нові напрями, як: плазмохімія, лазерна хімія, кріохімія, звукохімія та ін.

У повному обсязі ББК важлива тільки для бібліотекарів, що організовують і поповнюють фонди бібліотек. Ученому-хімікові

достатньо основних понять про цю систему.

16.3. Авторський знак

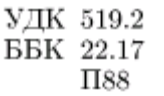
Авторський знак – один з основних елементів вихідних відомостей друкованого видання, який ввела відомий фахівець бібліотечної справи Любов Борисівна Хавкіна у 1916 р. Складається з букви та двох цифр. Буква – перша буква прізвища автора або назви книги. Цифри визначаються за спеціальними таблицями, в яких кожній послідовності з декількох перших літер прізвища автора чи назви книги (якщо авторами книги є більш ніж троє людей, або авторів немає) відповідає двозначне число. Останнє, 25-те видання таблиць Л. Б. Хавкіної "Таблица авторских знаков двоичных" вийшло у видавництві "Либерия" 1992 р. [27].

Приклад 35. *Визначення авторського знака за таблицями Хавкіної:*

Першим автором цього посібника є Кормош Ж. О. В таблицях знаходимо, що першим літерам прізвища "корм" відповідає код "К66".

Розташування авторського знака в друкованому виданні у СРСР визначалося за ГОСТ 7.4-77 та ГОСТ 7.4-86, а в Україні його визначає ДСТУ 4861:2007. Для книжкових видань це – верхній лівий кут обороту титульного листа – відразу під індексом ББК.

Приклад 36. *Авторський знак у друкованому виданні:*



В американській книжковій класифікації є "Кеттерівські таблиці" (cutter table) з елементами такої ж структури, але буква та двозначне число в них позначають рубрику й категорію, до якої належить книга, а не автора чи назву.

16.4. CAS

CAS registry number (він же CAS number, CAS RN, CAS #) – унікальний цифровий ідентифікатор хімічних сполук, біологічних послідовностей нуклеотидів та амінокислот, сумішей, внесених до реєстру Chemical Abstracts Service.

Номер CAS записують у вигляді трьох арабських чисел, розділених дефісами.

Хімічна реферативна служба (англ. Chemical Abstracts Service) – підрозділ Американского хімічного товариства (American Chemical Society) привласнює цей ідентифікатор усім речовинам, які хоча б один раз згадуються в літературі. Унікальний ідентифікатор призначений для більш зручного пошуку згадок у літературі завдяки усуненню проблеми різного найменування одного і того самого елемента. Нині практично усі бази даних з хімії мають пошукову систему за реєстраційним номером CAS.

CAS власне підтримує і продає базу даних хімічних речовин – реєстр CAS (англ. CAS regisry). На 15 січня 2010 р. у ньому містилося 51 804 164 органічних та неорганічних сполук, і 61 542 056 нуклеотидних та амінокислотних ланцюжків. Кожного тижня в базу додається приблизно 50 тис. нових записів. Актуальний розмір бази можна подивитися за посиланням: http://www.cas.org/cgi-bin/cas/regreport.pl.

Реєстраційний номер CAS є послідовністю цифр, розділеною знаками дефісу на три секції; перша частина може містити до 7 цифр, друга – дві цифри, третя складається з однієї цифри і виконує функцію контрольного символа. Номери призначаються у порядку зростання і не мають заздалегідь визначеного значення. Контрольна сума обчислюється через складання останньої цифри номера, помноженої на 1, другої справа цифри, помноженої на 2, третьої, помноженої на три і так далі до першої зліва цифри, завершуючись обчисленням залишку від ділення на 10.

Приклад 37. *Розрахунок контрольної суми реєстраційного номера CAS для води: 7732-18-5:*

$$8 \cdot 1 + 1 \cdot 2 + 2 \cdot 3 + 3 \cdot 4 + 7 \cdot 5 + 7 \cdot 6 = 105$$

$$105 \bmod 10 = 5 .$$

Окремі ізомери молекул також отримують власний номер CAS.

Приклад 38. *CAS# ізомерів глюкози:*

D-глюкоза –50-99-7;
L-глюкоза – 921-60-8;
β-D-глюкоза – 26655-34-5.

Іноді єдиний номер CAS призначають для усього окремого класу: група алкогольдегідрогеназ має номер 9031-72-5.

Прикладом, коли номер визначений для суміші, може слугувати гірчичне масло (8007-40-7) – за цим номером мається на увазі натуральне ефірне масло гірчиці, що складається, проте, майже з чистого аллілтіоізоціанату (CAS 57-06-7).

Використовуючи номери CAS для пошуку в базах даних, у запиті корисно вказувати і номери близькоспоріднених сполук. Наприклад, для пошуку інформації про кокаїн (CAS 50-36-2), шукати потрібно і кокаїну гідрохлорид (CAS 53-21-4), оскільки як наркотичний засіб поширеніше використання саме цієї форми сполуки.

Монополія CAS, яка склалася за останні роки, на надання платної послуги на привласнення порядкового номера, що не несе ніякої хімічної інформації, у багатьох випадках ускладнює його використання як код/покажчик для звернення до інформації.

16.5. PURL

PURL (англ. Persistent Uniform Resource Locator) – постійне загальноприйняте місцерозташування. Формат PURL заснований на URL – стандартизованому способі запису адреси ресурсу в мережі Інтернет. PURL, на відміну від звичайного URL, не змінюється при зміні Інтернет-адреси документа.

Проект PURL був розроблений та запущений у січні 1996 р. і нині підтримується організацією OCLC (с. 158).

Система серверів PURL подібна до системи серверів DNS, але, на відміну від останнього, субдомени в PURL безкоштовні. Йдеться про те, що якщо не сплатити за домен DNS, то він буде видалений із бази або переєрестрований на іншу особу. А ось постійний локатор PURL існуватиме поки не зникне сам проект PURL, а це, мабуть, відбудеться нескоро (OCLC існує досить довго, її закриття не передбачається).

Систему PURL у світі використовують сотні бібліотек, каталогів, словників, довідників, енциклопедій та інших ресурсів. Вона призначена для наукових і дослідницьких організацій, некомерційних фондів і суспільних асоціацій, а також для приватних осіб.

За своєю функціональністю PURL – це той же URL, але він указує не на конкретне місце розміщення ресурсу, а на запис у базі даних PURL, де, у свою чергу, записана вже конкретна URL-адреса ресурсу. При зверненні до PURL сервер знаходить потрібний запис у своїй базі даних і перенаправляє запит вже на конкретне місцерозташування ресурсу. Якщо адреса ресурсу змінюється, то немає потреби виправляти всі незліченні посилання на нього – достатньо лише змінити запис (конкретну адресу) в БД, а PURL залишиться абсолютно незмінним. Так досягається постійність зовнішньої адреси ресурсу. Структура адреси PURL аналогічна зі структурою звичайної веб-адреси. По суті, вона і є веб-адресою.

Приклад 39. *Структура PURL:*

протокол://сервер PURL/ім'я.

Приклад 40. *Приклад PURL:*

PURL адреса: http://purl.access.gpo.gov/GPO/LPS57324*;*
реальна веб-адреса, котра може змінюватись: http://arq.lanl.gov.

Слід зазначити, що підтримкою PURL у всіх видавництвах займається особа, зацікавлена в його збереженні. Тому можна припустити, що з часом буде зростати кількість "мертвих" PURL-адрес, котрі не ведуть до шуканого електронного документа.

16.6. SMILES

SMILES (англ. Simplified Molecular Input Line Entry Specification – специфікація входу спрощеного молекулярного ввідного рядка) – система правил (специфікація) однозначного опису складу і структури молекули хімічної речовини з використанням рядка символів ASCII. Назва в англійській мові є омонімом до слова smiles (посмішки), проте пишеться тільки

великими буквами. В українській мові однозначного аналога немає, рекомендоване вживання мовою оригіналу.

На основі рядка символів, складеного за правилами SMILES, більшість хімічних редакторів можуть побудувати двовимірну або тривимірну структурну формулу молекули.

Первинний варіант специфікації SMILES розробили Артур Вейнінгер (Arthur Weininger) і Девід Вейнінгер (David Weininger) в кінці 1980-х рр. Надалі стандарт був модифікований і розширений завдяки активному сприянню компанії Daylight Chemical Information Systems, Inc.

Головною перевагою SMILES над іншими ідентифікаторами хімічних сполук є, зокрема, краще сприйняття формул людиною, а також простіша програмна підтримка завдяки з наявності обширної теоретичної бази – теорії графів.

16.6.1. Варіанти специфікації SMILES

У початковій специфікації SMILES немає правил, що стосуються способу побудови запису і способу розрізнення просторових ізомерів молекул. Для вирішення цих проблем розроблено розширення стандарту.

Канонічна SMILES (Canonical SMILES) – версія специфікації, включно з правилами канонізації, що дають змогу записати формулу молекули будь-якої речовини в однозначному варіанті. Ці правила стосуються вибору першого атома в записі, напряму обходу циклів, вибору напряму основного ланцюга при розгалуженнях. Оскільки в різних пакетах молекулярного моделювання використовують різні алгоритми канонізації SMILES, унаслідок чого можна створити різні записи однієї і тієї ж молекули, поняття канонічна SMILES не є абсолютним. Цю версію стандарту застосовують зазвичай для індексування і перевірки унікальності молекул у базах даних.

Ізомерна SMILES (Isomeric SMILES) – версія специфікації, що дає змогу включати в запис дані про ізотопний склад, конфігурацію асиметричних атомів вуглецю і подвійних зв’язків. Особливість цієї версії порівняно з офіційною номенклатурою ІЮПАК у тому, що ізомерна SMILES дає змогу зберігати інформацію про молекули, для яких відомі конфігурації лише деяких хіральних центрів або подвійних зв’язків.

16.6.2. Визначення в термінах теорії графів

У термінах теорії графів SMILES являє собою рядок, отриманий через виведення символів вершин молекулярного графа внаслідок обходу їх у глибину молекули. Первинна обробка графа передбачає видалення атомів водню і розбиття циклів так, щоб граф, що вийшов, являв собою остовий ліс. На місцях розбиття графа ставляться цифри, що показують наявність зв'язку в початковій молекулі. Для позначення точок розгалуження молекули використовують дужки.

16.6.3. Основні принципи побудови SMILES

16.6.3.1. Атоми

Атоми позначають символами хімічних елементів у квадратних дужках, наприклад золото – як [Au]. Для елементів-органогенів (B, C, N, O, P, S, F, Cl, Br, I) дужки можуть бути пропущені. У цьому разі атоми водню можна не вказувати у явному вигляді, якщо їх кількість відповідає найменшій нормальній валентності відповідно до заданих зв'язків. Атоми у складі ароматичних циклів записуються малими літерами замість великих, хоча у деяких діалектах застосовують "кекулізацію" SMILES з чергуванням подвійних та одинарних зв'язків. При потребі вказати формальний заряд частинки атоми водню і символ заряду записують у вигляді. Ізотопи вказують у квадратних дужках із зазначенням атомної ваги перед символом атома, наприклад, ізотоп ^{13}C буде записаний як [13C].

Наприклад, запис SMILES для води виглядатиме як O, для етанолу – CCO. Гідроксильний аніон записується [OH-], а іон заліза (II) – [Fe+2].

16.6.3.2. Зв'язки

Одинарний хімічний зв'язок може бути записаний із використанням символу "-" між атомами, сполученими зв'язком, але на практиці дефіс пропускають. Позначення ароматичного зв'язку (:) зазвичай теж не використовують. Подвійний зв'язок позначається за допомогою знака рівності, наприклад, двоокис вуглецю записують як O=C=O. Потрійний зв'язок позначається за допомогою символу решітки "#", наприклад, синильна кислота записується як C#N.

16.6.3.3. Розгалуження молекули

Бокові ланцюги молекули беруть у круглі дужки. Наприклад, пропіонова кислота записується як CCC(=O)O. Канонічна форма запису трифторметану виглядає як C(F)(F)F, проте такий запис незручний для читання через свою перевантаженість дужками, тому ту ж молекулу можна записати в неканонічній формі як FC(F)F.

16.6.3.4. Циклічні сполуки

Атоми, що містяться на кінцях розірваного при побудові основного розгалуження зв'язку, позначають одним і тим же номером. Наприклад, циклогексан записується як C1CCCCC1, а бензол – як c1ccccc1 (зверніть увагу на регістр букв для запису циклогексану та бензолу).

16.6.3.5. Стереохімія

Конфігурація щодо подвійного зв'язку записується за допомогою символів / і \. Наприклад, F/C=C/F відповідає транс-дифторэтилену, а F/C=C\F або F\C=C/F відповідає цис-дифторэтилену.

16.6.3.6. Розширення

SMARTS – модифікація SMILES, яка дає змогу використовувати невпорядковану структуру атомів і зв'язків, їх широко використовують в системах пошуку в базах даних про речовини. Практика застосування викликала поширену хибну думку, що в комп'ютерному пошукові структур проводиться порівняння записів-ланцюжків, тоді як насправді проводиться набагато ефективніше порівняння графів, побудованих на підставі формул SMILES.

16.6.3.7. Перетворення

Формула SMILES може бути перетворена у двомірну структурну формулу за допомогою алгоритму (Structure Diagram Generation algorithms), який розробив Херсон, і не завжди дає однозначний результат. Перетворення у тривимірну структурну формулу проводиться з використанням принципу мінімальної

енергії утворення речовини.

Приклад 41. *Приклад SMILES молекули диклофенаку:*

c1ccc(c(c1)CC(=O)O)Nc2c(cccc2Cl)Cl .

Приклад 42. *Приклад SMILES молекули цимантрену:*

[CH]1[CH][CH][CH][CH]1.C(#O)[Mn](C#O)C#O .

16.7. Комплексний книготорговий індекс-шифр

Комплексний книготорговий індекс-шифр – елемент вихідних відомостей, призначений для класифікації книги у книготорговій мережі (книжкових магазинах і т. д.).

Комплексний книготорговий індекс-шифр почали використовувати у книгах ще в СРСР 1 липня 1965 р. разом із введенням в дію "Єдиної схеми класифікації літератури в книготорговій мережі" (рос. ЕСКК – Единая схема классификации литературы в книготорговой сети).

У цьому виданні формат індексу-шифру був визначений як дріб, у чисельнику якого – класифікаційний індекс (набір яких визначено там же), а в знаменнику указували видання, у якому оголошували про майбутній випуск книги. Класифікаційний індекс складався з трьох одно- або двозначних чисел, розділених дефісами. Перше число позначало відділ класифікації, друге – підвідділ, а третє – ще дрібніше ділення – групу. Згодом дефіси замінили нулями, причому так, що індекс став п'ятизначним. У знаменнику вказувався номер позиції в каталозі видання, його назва у вигляді абревіатури, номер цього видань і рік його випуску. За замовчуванням – якщо абревіатура не указувалася – як каталог видань мали на увазі тематичний план цього видавництва з випуску літератури. Видання, у яких оголошували про план видавництва з випуску літератури на майбутній рік так звані тематичні плани, були призначені, зокрема, для відвідувачів книжкових магазинів, які, проглянувши плани видавництв, могли заздалегідь залишити в магазині замовлення на книги.

У 1977 р. видана "Єдина схема класифікації літератури для книговидання в СРСР", у якій комплексний книготорговий індекс-шифр зазнав істотних змін (розділ "Структура"), але в багатьох книгах, виданих наступного року, використовували п'ятизначний індекс ЕСКК. Тому період 1979–1981 рр. був визначений як

перехідний, коли в комплексному книготорговому індексі-шифрі книг, виданих у СРСР, повинен був указуватися як п'ятизначний, так і десятизначний класифікаційний індекси. У 1981 р. п'ятизначний індекс був остаточно скасований, і з 1982 р. комплексний книготорговий індекс-шифр у радянських книгах мав формат, визначений у виданні "Єдиної схеми" 1977 р. та закріплений в ГОСТ 7.4–86. У вихідних відомостях видань він використовувався до набуття чинності ГОСТ 7.4–95.

16.7.1.1. Структура

Комплексний книготорговий індекс-шифр мав структуру:

$$A\frac{B-C}{D(E)-G}H,$$

де:

A – перша буква авторського знака; B – десятизначний код "Єдиної класифікації літератури для книговидання в СРСР"; дві перші цифри – це розділ тематичної класифікації; наступні дві – підрозділ; далі – рубрика; потім – підрубрика; останні дві цифри – ще дрібніший поділ. Відповідні коди опубліковані в "Єдиній класифікації". Згідно з ГОСТ 7.4–86 у книгах, брошурах і нотних виданнях десятизначний код розташований у лівому нижньому кутку відвороту титульного листа, а в однолистних виданнях – у лівому нижньому кутку поля листа; C – номер видання серед випущених певним видавництвом протягом звітного року; D – код видавництва згідно з "Єдиною класифікацією літератури для книговидання в СРСР"; E – код відомства, якому підпорядковане видавництво; G – рік випуску видання; H – документ, у якому було оголошено про вихід цього видання, або "без об'явл."

Приклад 43. *Книготорговий шифр-індекс практикуму з технології напівпровідників:*

Книга "Практикум по химии и технологии полупроводников: учеб. пособие для студентов вузов / В. З. Анохин, Е. Г. Гончаров, Е. П. Кострюкова и др.; Под ред. Угая Я. А. – М. : Высш. шк., 1978. – 191 с."

$$П\ \frac{20507—340}{001(01)—78}\ 56—78.$$

Приклад 44. *Книготорговий шифр-індекс підручника з хімічної технології:*

Книга "Бесков В. С. Общая химическая технология и основы промышленной экологии: учебник для вузов / В. С. Бесков, В. С. Сафронов. – М. : Химия, 1999, 472 с. – ISBN 5-7245-1133-9" має такий індекс:

$$Б\ \frac{2802000000—006}{050(01)—99}\ Без\ объявл.$$

Сьогодні книготорговий шифр-індекс має історичне значення, хоча і трапляється в багатьох книжках, що стали вже класикою, але не втратили свою актуальність. Його змінили ISSN.

16.8. ISSN

ISSN (англ. International Standard Serial Number – міжнародний стандартний серійний номер) – унікальний номер, що дає змогу ідентифікувати будь-яке серійне видання незалежно від місця видання, мови та фізичного вигляду.

У 1975 р. введено стандарт ISO 3297, що визначає правила присвоєння ISSN. За сприяння ЮНЕСКО та уряду Франції видача номерів здійснюється у 75-ти національних центрах.

Цей стандартний серійний номер широко використовують у всьому світі: він потрібний бібліотекам, підписним агентствам, дослідникам і вченим, агентствам новин і т. п. Мережа організацій ISSN втілює широкомасштабну програму міжнародної співпраці. Вона вже охоплює 239 країн і безперервно розширюється.

Присвоєння номерів ISSN в Україні здійснюється згідно зі стандартом ГОСТ 7.56–2002 (ISO 3297-98) – міжнародна стандартною нумерацією серійних видань (International Standard Serial Numbering ISSN), що регламентує правила надання ISSN, котрий уведений в дію з 1 січня 2003 р. Складається з восьми цифр: восьма цифра – контрольное число, що розраховується за попередніми сімома і модулем 11.

ISSN дає змогу видавцям, книгопродавцям, бібліотекарям, науковим співробітникам визнаним у всьому світі способом безперешкодно здійснювати розповсюдження серійних видань залежно від попиту, удосконалити пошук і замовлення видань, весь цикл створення і доведення серійного видання до споживача.

ISSN є унікальним ідентифікатором для кожного серійного видання, обов'язковим елементом вихідних відомостей серійних видань. На основі цифрового ISSN створюють штриховий код для серійних видань.

Система ISSN застосовується більше ніж у 130 країнах. Використання цифрового коду, відображеного на виданні, дає можливість відмовитися від локальних кодів, скоротити кількість супровідної документації, спростити взаємодію видавництв з поліграфічними підприємствами, розповсюджувачами і бібліотеками, шукати інформацією про видання в автоматизованих системах за допомогою ISSN на національних і міжнародних рівнях заощаджувати кошти, зусилля і т. д. Блоки ISSN встановлює і розподіляє міжнародний центр ISSN.

Це позначення присвоюють будь-якому періодичному виданню (публікації з будь-якої галузі знань, у послідовно випущених частинах, що мають числові або хронологічні позначення і періодичний характер). Це може бути газета, тижневик, журнал, щорічник, а також електронна публікація (CD-ROM, веб-сайт). Фундаментальним критерієм є періодичність, тобто частини публікації видають як відкритий набір документів з одним і тим же заголовком, терміном і на певний час, не обмежений заздалегідь. ISSN використовують для виготовлення штрихового коду ідентифікації, що використовують у всьому світі. Як стандартний числовий код ідентифікації ISSN придатний для комп'ютерного застосування, пошуку і передачі даних. Як читаний код ISSN є точним цитуванням для публікації, його використовують дослідники, учені й бібліотекарі. ISSN – фундаментальний інструмент для ефективного постачання документів, що забезпечує корисний та економічний метод зв'язку між видавцями і постачальниками, роблячи торгові системи розповсюдження ефективнішими.

Якщо наявні різні версії однієї і тієї ж публікації на різних носіях (друковане видання, CD-ROM, Інтернет), то кожна версія

повинна отримати окремий номер ISSN, навіть якщо зміст публікацій ідентичний.

Процедура надання та використання ISSN регулюється регламентом Міжнародного центру ISSN і Національними центрами. В Україні – згідно з ГОСТ 7.56-89.

16.8.1. ГОСТ 7.56-89

Чинний стандарт розповсюджується на спосіб ідентифікації серійних видань на основі застосування міжнародної стандартної нумерації серійних видань (International Standard Serial Numbering – ISSN) і встановлює структуру, форму уявлення, розміщення у виданнях і процедуру присвоєння ISSN, утворення ключового заголовка і набір елементів даних, потрібних для ідентифікації серійних видань. До серійних видань входять такі, що випускаються протягом часу, тривалість якого заздалегідь не встановлена, виходять послідовно через певні або невизначені проміжки часу, однотипно оформленими нумерованими і (або) датованими випусками, що мають загальний заголовок і зазвичай однаковий формат.

Основними серійними виданнями є:

1) газети;
2) журнали;
3) щорічники;
4) серії;
5) доповіді, звіти, збірки статей, що публікують періодично;
6) праці, бюлетені, що виходять періодично;
7) офіційні видання конгресів і конференцій, що проводяться періодично.

Стандарт обов'язковий для всіх видавництв і організацій, що випускають серійні видання, незалежно від їх відомчого підпорядкування.

16.8.2. СТРУКТУРА ISSN

Міжнародний стандартний номер серійних видань складається з абревіатури ISSN і подальших восьми цифр. ISSN складається з двох чотиризначних цифрових груп, що розділяються дефісом. Абревіатура ISSN і перша цифра розділяються пропуском. Для ISSN застосовують арабські цифри від 0 до 9. Остання цифра ISSN (контрольна) може бути римською цифрою X, що використовується для позначення числа 10. Контрольна

цифра слугує для перевірки правильності написання цифрової частини ISSN.

ISSN використовується як реєстраційний (порядковий) номер.

16.8.3. НАДАННЯ ISSN

Блоки ISSN встановлює і розподіляє Міжнародний центр (МЦ), відповідальний за реєстрацію серійних видань, що виходять у світі.

Кожному серійному виданню надають тільки один ISSN. Один і той же ISSN не може бути наданий двічі.

За надання ISSN не виплачується авторський гонорар. На використання ISSN і ключового заголовка авторського права немає.

16.8.4. КОДИ, ЯКІ ЗАЗНАЧАЮТЬ РАЗОМ З ISSN

Разом з ISSN може зазначатися міжнародний стандартний номер книги ISBN (International Standard Book Number).

Для частин (томів, випусків) із окремими заголовками (монографічного характеру), яким надано ISSN, паралельно надається і ISBN – для кожного тому окремо згідно з ГОСТ 7.53.

16.8.5. РОЗМІЩЕННЯ ISSN

ISSN наводять за ГОСТом 7.4 і поміщають у верхньому правому кутку першої сторінки обкладинки або палітурки, якщо цих сторінок немає – на титульній сторінці або суміщеній титульній сторінці кожного випуску серійного видання. Якщо ISSN наданий для загального заголовка і заголовка серії (підсерії), то спочатку подають ISSN загального заголовка, а під ним – ISSN заголовка серії (підсерії).

ISSN входить до складу бібліографічного рядка, поміщеного вертикально уздовж корінця видання або на четвертій сторінці обкладинки чи палітурки в окремо відведеному місці.

Якщо серійному виданню надано два ISSN, обидва номери зазначають разом з абревіатурами на звороті титульної сторінки при збереженні ISSN на першій сторінці обкладинки або на титульній сторінці. Рекомендується проставляти ISSN у лівому нижньому кутку звороту титульної сторінки під ISBN.

Якщо в межах серійного видання випускається додаток із самостійною титульною сторінкою, то на основному виданні ставиться основний ISSN, а на додатку – свій ISSN.

ISSN розміщують у складі вихідних відомостей у верхній

частині першої колонки або в нижній частині останньої колонки кожного випуску газети.

Шрифт ISSN повинен відрізнятися від шрифту основного тексту видання.

16.8.6. ISSN І КЛЮЧОВИЙ ЗАГОЛОВОК СЕРІЙНОГО ВИДАННЯ

Ключовий заголовок – назва, встановлена для ідентифікації та реєстрації певного серійного видання у процесі надання ISSN.

Ключовий заголовок встановлюють на основі вихідних відомостей про видання (назви видання, вихідних даних і т. д.).

Кожному ISSN відповідає окремий ключовий заголовок і кожному ключовому заголовку відповідає один ISSN.

При зміні заголовка видання міняється ключовий заголовок і ISSN.

Якщо в межах серійного видання випускають додаток із самостійною титульною сторінкою, обом серійним виданням встановлюють окремі ключові заголовки й ISSN.

Приклад 45. *Приклад ISSN:*

ISSN 0321-4095 (Науково-технічний журнал "Вопросы химии и химической технологи" / ДВНЗ "Український державний хіміко-технологічний університет", м. Дніпропетровськ)

16.9. ISBN – Міжнародний стандартний номер книги

ISBN (англ. The International Standard Book Number, міжнародний стандартний номер книги) – універсальний ідентифікаційний код, який проставляють на книгах і брошурах незалежно від способу їх виготовлення, розповсюдження, тиражу та обсягу.

ISBN однозначно й безпомилково ідентифікує лише одне неперіодичне видання одного видавця, є неповторним і використовується тільки для цього видання. ISBN є ключем для пошуку потрібних видань, які публікують у світі, в автоматизованих системах на національному та міжнародному рівнях. Використання ISBN дає змогу об'єднати в єдину систему видання, книгорозповсюдження та інформаційне обслуговування.

Стандарт розроблено у Великобританії 1966 р. на базі

9-значного Стандартного номера книг (Standard Book Numbering (SBN) code) Гордона Фостера (Gordon Foster). У 1970 р. із незначними змінами був прийнятий як міжнародний стандарт ISO 2108. Номери ISBN, привласнені книгам до 2006 р. видання включно, складаються з 10 символів. З 2007 р. набрав чинності новий стандарт ISBN – 13-значний, котрий збігається зі штрихкодом книги. На території колишнього Радянського Союзу ISBN почали використовувати з 1987 р.

Міжнародний стандартний номер книги складається з абревіатури ISBN (незалежно від мови видання) й тринадцяти цифр. Для позначення цифрової частини ISBN застосовують арабські цифри від 0 до 9. Цифрова частина ISBN складається з п'яти груп цифр, три з яких містять різну кількість цифрових знаків, відокремлених одна від одної дефісом або проміжком, а перша й остання група – фіксовану (сталу). Цифрова частина відокремлюється від абревіатури ISBN проміжком. П'ять груп цифр ISBN розміщені в такій послідовності:

– префікс GS1(EAN·UCC) для книг (978, до 1 січня 2007 р. не використовувався);

– ідентифікатор групи (країни) залежить від мови видання. Ідентифікатори України – 966 та 617;

– ідентифікатор видавця;

– порядковий ідентифікатор видання;

– контрольна цифра.

Число цифр в ідентифікаторі групи залежить від обсягів випуску книжкової продукції (може бути більше однієї), наприклад: 0 і 1 – група англомовних країн, 2 – франкомовних, 3 – німецькомовних, 4 – японськомовних, 5 – російськомовні країни (деякі країни колишнього СРСР, Росія), 7 – китайська мова, 80 – для Чехії та Словаччини, 600 – Ірану, 953 – Хорватії, 985 – Білорусі, 9956 – Камеруну, 99948 – Еритреї. Загалом групам надано номери 0–7, 80–94, 950–993, 9940–9989, і 99900–99999.

Код видавництва присвоює Національне агентство ISBN, при цьому беруть до уваги кількість публікацій, яку видавець має намір здійснити. Потужнішим видавництвам надають коротший номер, щоб зробити доступним більше знаків для нумерації видань (сумарний обсяг номерів видавця і видання

для ISBN, що надає українське агентство, становить шість цифр). Можливе також надання унікального номера видання (в Україні – від 10 до 12 знаків). Контрольна цифра (арабська від 0 до 9 або римська X) слугує для перевірки правильності числової частини ISBN. Розрахунок здійснює Національне агентство ISBN.

Для використання як штрихкодів формату EAN-13 до ISBN додають префікс GS1 978, наданий Асоціацією GS1, і замість контрольної цифри ISBN використовується контрольна цифра, розрахована згідно зі стандартом EAN-13.

Приклад 46. *ISBN книги згідно з форматом EAN-13:*

ISBN 978-966-8602-34-4 має такі складники:
978 – префікс для книг;
966 – ідентифікатор України;
8602 – ідентифікатор видавця;
34 – порядковий ідентифікатор видання;
4 – контрольна цифра.

На основі ISBN будується штриховий код символіки EAN-13 для книжкового видання.

Ідентифікатори для видань присвоюють Національні агентства з міжнародної стандартної нумерації книг. В Україні це – Книжкова палата України імені Івана Федорова. На початок 2011 р. вартість послуги становила 33 грн 70 коп. без податку на додану вартість.

Для видань, що виходять малим тиражем, або для "особистого" використання надавати номер ISBN необов'язково.

Приєднання до системи ISBN накладає на видавця відповідальність:

– за використання тільки тих ISBN, які надало Національне агентство ISBN;

– за надання, розміщення, форму зазначення ISBN у виданні;

– за несанкціоноване використання ISBN іншого видавництва для своїх видань;

– за несанкціоновану передачу ISBN іншому видавництву;

– за інформування Національного агентства ISBN про використані номери, про зміну назви, юридичної адреси,

припинення діяльності;

– за своєчасне отримання нового номера регістранта у разі зміни назви видавництва та в інших випадках, передбачених у системі ISBN.

ISBN є обов'язковим елементом вихідних даних. Його поміщають у нижньому лівому кутку обороту титульної сторінки або в нижній лівій частині суміщеної титульної сторінки. Кожна нова книга, кожне її перевидання, переклад іншою мовою або випуск із новим оформленням повинні мати власний міжнародний стандартний номер.

На виданні можуть стояти два і більше номери ISBN, якщо це:

– багатотомне видання (номер тому і номер видання);

– сумісне видання (номери кожного видавця із вказаними у круглих дужках їх найменуваннями після відповідного ISBN);

– переклад видання, опублікований вперше (номер перекладу і номер оригіналу із зазначенням у круглих дужках відомостей про мову після відповідного ISBN);

– комплектне видання, тобто зібране в теку, футляр або поміщене в загальній обкладинці (власний номер книги і ISBN, загальний для всього комплекту).

ISBN дає змогу вести оперативний пошук інформації про конкретне видання в різних інформаційних ресурсах, удосконалювати замовлення книг, здійснювати контроль за їх продажем.

Відомості про видавця (назви, ідентифікатори ISBN, адресні дані, спеціалізація) передають в Міжнародне агентство ISBN для випуску Міжнародного покажчика видавництв і видавничих організацій (Publishers' International ISBN Directory).

Міжнародна стандартна нумерація книг не розповсюджується на:

– періодичні видання, що продовжуються (серійні), (журнали, газети, бюлетені, збірки, що періодично продовжуються, пронумеровані щорічники);

– призначені для тимчасового використання друкарські видання (рекламні, роздаткові матеріали, програми заходів, календарні плани, календарі, що не є виданнями книжкового типу, товаросупровідні документи);

– видання з будь-якими обмежувальними позначками;

– листові видання;

– нотні видання;

– образотворчі видання;

– картографічні видання (окрім атласів);

– автореферати дисертацій;

– препринти;

– окремі видання нормативно-технічних документів (патенти, стандарти, прейскуранти);

– конспекти лекцій, учбові програми і плани, видання в картковій формі.

16.10. Цифровий ідентифікатор об'єкта (DOI)

DOI (англ. Digital Object Identifier – цифровий ідентифікатор об'єкта) – унікальний ідентифікатор об'єктів інтелектуальної власності. Він може існувати в електронній, фізичній чи абстрактній формі, яка залежить від системи, що оперує цими об'єктами. Індекс DOI зазвчай не пов'язаний зі змістом чи певними властивостями об'єкта, але головною метою його створення була можливість оперування індексами DOI через Інтернет, а призначення – ідентифікація посилань.

Загальне керівництво системою DOI здійснює International DOI Foundation Inc. (IDF), котра була заснована 1998 р., а обліком інтелектуальних об'єктів займаються реєстраційні агентства. У галузі науки реєстраційним агентством є організація CrossRef, створена провідними міжнародними й національними видавництвами.

Сьогодні CrossRef реєструє наукові статті, книги й розділи книг; цілком імовірно, що у майбутньому об'єктами реєстрації стануть і фрагменти статей, наприклад, ілюстрації. Імена DOI можуть ідентифікувати журнал, індивідуальний результат журналу, окрему статтю в журналі або єдину таблицю у цій статті. Вибір ступеня деталізації залежить від того, хто призначає цей номер, але в системі DOI це повинно бути зазначено як частина супровідних метаданих.

Нині основні галузі використання DOI такі:

– постійні цитати в академічних матеріалах (статті журналів, книги і т. п.) через консорціум близько 3000 видавців;

– наукові набори даних через консорціум найбільших бібліотек, технічні й наукові інформаційні центри;

– офіційні публікації Євросоюзу, через центри публікацій Євросоюзу.

Ім'я DOI не залежить від розміщення об'єкта, тому подібне до Uniform Resource Name (URN) або Persistent Uniform Resource Locator (PURL), але відрізняється від Uniform Resource Locator (URL).

Надання імені DOI платне. Суми грошових зборів встановлює незалежно кожне Реєстраційне агентство окремо. Цими коштами компенсують витрати на забезпечення функціонування системи. Своєю чергою адміністративний грошовий збір РА платить до IDF, щоб забезпечити витрати на розвиток і підтримку системи.

Зацікавлене видавництво стає членом CrossRef і отримує право формулювати DOI для своєї наукової продукції. Видавництво бере на себе зобов'язання надсилати в CrossRef для внесення в централізовану базу даних такі відомості про кожну статтю: код DOI, бібліографічний опис і URL тієї сторінки, на якій міститься або власне стаття, або її бібліографічний опис і правила доступу до повного тексту (в разі платного доступу). Код DOI повинен бути зазначений на сторінках відповідної статті, і будь-який читач може звернутися до бази даних CrossRef для того, щоб отримати поточний URL цієї публікації.

Видавництво зобов'язане повідомляти в CrossRef про будь-яку зміну адреси статті, при цьому до бази даних вносять виправлений URL, але DOI залишається таким же. DOI не змінюється, якщо журнал переходить до іншого власника або змінюється статус статті (що пройшов рецензування препринт → онлайнова публікація → → публікація в друкованому виданні).

Код DOI обов'язково починається з числа 10 і має таку структуру: 10, знак крапки ".", код видавництва, знак "/", суфікс – це унікальний в межах кокретного видавництва набір символів, однозначно належить саме до цієї статті, наприклад:

10.1000/182,

де10.1000 – префікс, 10 – код каталогу, який має позначатися цифрою; коректні коди каталогів визначають агенції з підтримки (IDF). Сьогодні єдиним кодом каталогу є 10, і всі коректні DOI починаються з "10.";

1000 – код реєстратора або іденитфікатор видавця, який ідентифікує реєстратора. У нашому прикладі код ідентифікує IDF;

182 – суфікс, або ідентифікатор екземпляра, який ідентифікує окремий об'єкт (звичайні суфікси більші за обсягом порівняно з наведеним прикладом). Тоді об'єкт, що відповідає doi:10.1000/182, є "DOI Handbook, Version 4.4.1".

Префікс надає Агенція з реєстрації DOI (англ. DOI Registration Agency) безпосереднім реєстраторам. Суфікс надається реєстратором і має бути унікальним у контексті цього префікса. До нього можуть входити стандартні ідентифікатори, такі як ISBN або ISSN. DOI не залежить від регістра літер.

Цитувати DOI на веб-сторінці або в публікації слід так: doi:10.1000/182.

Приклад 47. *Приклад коду DOI:*

doi:10.1016/j.msec.2008.09.003.

Передбачається, що в майбутньому в науковій сфері DOI застосовуватиметься замість URL; поки ж DOI виконує допоміжні функції.

CrossRef рекомендує вказувати код DOI скрізь, де подається бібліографічний опис електронного документа, причому не лише на веб-сторінках, а й у друкарських варіантах статей. Наприклад:

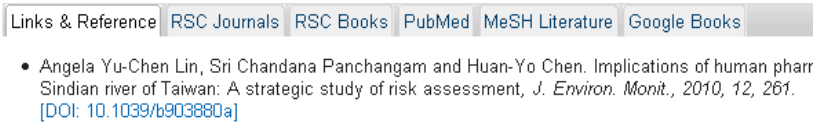


Рис. 17. *DOI в онлайновому змісті*

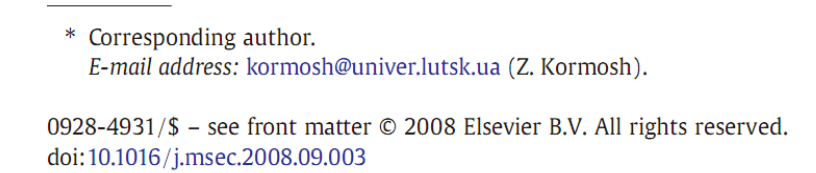


Рис.18. *DOI в статті у друкованому журналі*

Провідні видавництва використовують DOI і базу даних CrossRef для створення системи гіперзв'язків, яка об'єднує матеріали, опубліковані в різних журналах. Тоді у списку використаних літературних джерел подають не лише бібліографічні описи статей, а й посилання на конкретні онлайнові файли. На деяких сайтах у посиланні на статтю код DOI вказують у явній формі.

Система DOI реалізується через федерацію Реєстраційних агентств DOI, інфраструктуру забезпечує Міжнародний фонд DOI, який розширився і управляє системою. Система DOI була розвинена і започаткована у низці видавничих програм, починаючи з 2000 р. На квітень 2011 р. було надано більше ніж 51 млн імен DOI. До системи входить близько 4000 організацій.

16.11. Міжнародний хімічний ідентифікатор InChI

IUPAC InChI (англ. International Chemical Identifier – міжнародний хімічний ідентифікатор) – текстовий ідентифікатор хімічних речовин, створений для того, щоб забезпечити стандартний, зрозумілий людині спосіб кодування молекулярної інформації та полегшити пошук такої інформації в базах даних у мережі. InChI розробляли протягом 2000–2005 рр. агентства IUPAC і NIST. Його формат й алгоритми створення вільні для використання та удосконалення за ліцензією LGPL (хоча й ім'я "InChI" є торговою маркою IUPAC).

Цей ідентифікатор описує хімічні речовини за допомогою шарів інформації – атоми та їх зв'язки, таутомерна інформація, інформація про ізотопи, стереохімія й опис електричного заряду. Не обов'язково наводити всі шари інформації, наприклад, шар таутомерної інформації може бути пропущений, якщо вона не важлива у конкретному випадку.

Номер InChI суттєво відрізняється від поширеного реєстраційного номера CAS у трьох аспектах:

– він не є об'єктом інтелектуальної власності (на відміну від CAS), його можна вільно використовувати;

– він може бути розрахований на основі структурної інформації, жодна організація не займається його наданням;

– більшу частину інформації у номері InChI можна зрозуміти (звичайно, після певної практики).

Отже, InChI можна розглядати як загальну й дуже

формалізовану версію назви IUPAC.

Кожна структура має унікальний номер InChI, що є важливим при створенні та використанні баз даних інформації з хімії.

У номері InChI не міститься інформація про координати атомів у просторі, у такому разі використовують формати типу PDB.

На сьогодні можливість генерування номерів InChI на основі структурної інформації наявна в багатьох програмах з хімії.

Приклад 48. *InChI ацетону:*

InChI=1S/C3H6O/c1-3(2)4/h1-2H3

Приклад 49. *InChI піроксикаму:*

InChI=1S/C15H13N3O4S/c1-18-13(15(20)17-12-8-4-5-9-16-12)14(19)10-6-2-3-7-11(10)23(18,21)22/h2-9,19H,1H3,(H,16,17,20)

16.11.1. Структура ідентифікатора InChI

Кожен номер InChI починається з рядка "InChI=", потім вказується номер версії (на сьогодні це перша версія ідентифікатора – "1"). Подальша інформація структурована як послідовність шарів і субшарів, кожному з яких відповідає певний вид інформації. Ці шари та субшари відділені один від одного символом "/" і починаються з певної літери – префікса (за винятком субшару з хімічною формулою). Усього нараховують шість шарів із субшарами.

16.11.2. Шари InChI

Хімічна формула без префікса – це єдиний обов'язковий шар у кожному номері InChI.

Зв'язки між атомами (префікс "c"). Атоми у хімічній формулі (за винятком гідрогену) нумерують за порядком. Цей шар описує, які атоми і в якому порядку з'єднуються між собою. Додатково може містити:

– субшар пов'язаних між собою металів;
– субшар не пов'язаних між собою металів;
– ізотопи (префікс "i");
– стереохімію;
– подвійні зв'язки (Z/E);
– тетраедричну геометрію (sp^3);

– таутомерію.

Заряд (префікс “q”) вказується у кінці, хоча його не вважають частиною InChI.

16.11.3. Ключі InChI

Ключі InChI мають таке розшифрування:

/c – безпосередні зв’язки між атомами (без термінальних атомів гідрогену);

/h – зв’язки між атомами, включно з термінальними і лабільними атомами гідрогену;

/p – протонний баланс;

/t – тетраедральна геометрія;

/m – відносна стереохімічна конфігурація;

/s – стереохімічний тип (1 – абсолютна конфігурація, 2 – відносна, 3 – рацемат);

/i – ізотопи;

/f – хімічна формула разом з фіксованими атомами гідрогену (якщо відрізняється від брутто-формули);

/h – зв’язок між атомами (розміщення фіксованих лабільних атомів гідрогену);

/q – заряд;

/r – хімічна формула структури сполучених атомів металу.

Атоми гідрогену (префікс “h”) вказують на кількість атомів гідрогену, зв’язаних із кожним атомом молекули.

Шар заряду.

Субшар позитивного заряду (префікс “p”).

Субшар негативного заряду (префікс “q”).

Шар стереохімії.

Ізотопний шар.

Використання ключів сприяє більш зручному та ефективному пошукові – можна вказувати тільки потрібні поля певних шарів.

16.12. InChIKey

IUPAC InChIKey (англ. International Chemical Identifier Key – ключ міжнародного хімічного ідентифікатора) – хеш-версія InChI, створений для того, щоб полегшити роботу пошукових систем в Інтернеті. Він має стандартну довжину – 27 символів

(що значно менше, ніж у InChI) і легко індексується пошуковими машинами. Для людини він не несе жодної інформації про структуру молекули (крім кількості лабільних протонів). Перший стандарт InChIKey з'явився у 2007 р. (ключ складався з 25 символів). На початку 2009 р. започаткована версія 1.02 InChIKey (позначається Std. InChIKey, 27 символів), яку використовують і сьогодні.

InChIKey розраховують із повної версії InChI (використано алгоритм шифрування SHA-256). Через хеш-характер ключа InChIKey існує відмінна від нуля ймовірність, що дві різні сполуки одержать однаковий ключ InChIKey. Ймовірність збігу навіть у першій, 14-символьній частині, становить близько $1{,}3 \cdot 10^{-11}$. Вважають, що перший такий випадок станеться не раніше ніж через 100 років.

Стандартний ключ InChIKey має п'ять чітко окреслених компонентів:

AAAAAAAAAAAAAA – 14-симовольний хеш, розраховується на основі інформації про зв'язки між атомами та з атомами гідрогену;

BBBBBBBB – 8-символьний хеш, розраховується на основі інших шарів InChI-коду;

F – флаг, що вказує на стандартний InChIKey (розрахований зі стандартного InChI); на сьогодні має стале значення "S";

V – флаг, що вказує на версію InChI: "A" – для першої версії, "B" – для другої і т. д.

P – індикатор кількості протонів. Це не хеш, в кінці через дефіс (-) вказується літера, що відповідає надлишку чи нестачі протонів: "N" – для нейтральних, "M" – без одного протона, "O" – один надлишковий протон і т. д.

Отже, загальний обсяг ключа InChIKey становить 27 символів разом із розділовими дефісами:

AAAAAAAAAAAAAA-BBBBBBBBF-P

Цей запис значно коротший, ніж типовий рядок InChI.

Приклад 50. *InChIKey ацетону:*

CSCPPACGZOOCGX-UHFFFAOYSA-N

Приклад 51. *InChIKey піроксикаму:*

QYSPLQLAKJAUJT-UHFFFAOYSA-N

16.13. Питання для самоперевірки

1. Для чого потрібні класифікаційні індекси?
2. Чому існує потреба в хімічних ідентифікаторах?
3. Що таке УДК?
4. Принципи класифікації УДК.
5. Які розділи УДК використовують для хімічних наук?
6. Що таке ББК?
7. На чому грунтується класифікація ББК?
8. Що таке CAS registry number?
9. Що таке PURL?
10. Що таке SMILES?
11. Що таке “Комплексний книготорговий індекс-шифр”?
12. Що таке ISSN?
13. Структура ISSN.
14. Що таке ISBN?
15. Вкажіть структуру ISBN.
16. Що таке DOI?
17. InChI – це...
18. Структура InChI.
19. Що таке InChIKey?

17. ПОШУКОВІ МАШИНИ ЗАГАЛЬНОГО КОРИСТУВАННЯ

Першою глобальною пошуковою системою була AltaVista. Зараз її застосування різко зменшилося, але вона до цього часу зберігає восьме місце у рейтингу пошукових систем світу.

Найбільш популярними пошуковими системами у світі станом на квітень 2011 р. є Google (83,82 %) та Yahoo (5,88 %). Інші популярні в світі пошукові системи, такі як Bing (3,92 %) чи Baidu (4,38 %) у нас майже не відомі.

В Україні більш широко використовують регіональні пошукові системи Yandex та Rambler. Звичайно, є й національні, лише українські системи, але, зважаючи на малу корисність їх у пошуку наукової інформації з хімії, детально характеризувати їх не будемо.

Більш детально розглянемо найпоширенішу пошукову систему Google, потім – популярні регіональні Yandex та

Rambler. Торкнемося також англомовних пошукових систем AltaVista та Yahoo. У кінці розглянемо спеціалізовану пошукову систему наукової інформації Scirus.

17.1. Google

Пошукову систему Google (www.google.com) створили аспіранти Стенфордського університету Леррі Пейдж (Larry Page) і Сергій Брін (Sergey Brin) як навчальний проект. Вони з 1995 р. працювали над пошуковою системою BackRub, котра 15 вересня 1997 р. одержала назву "Google" (дещо спотворений термін "googol", який означав число, складене з одиниці й ста нулів, коли аспірантам вказали на помилку, назва була вже зареєстрована).

У 1998 р. для підтримки пошукової системи була створена компанія Google Inc., котра зараз є однією з найбільш перспективних компаній світу.

Google (вимовляється "гул") – перша за популярністю (83,82 % станом на квітень 2011 р.) у світі пошукова система, що обробляє більше 40 млрд запитів за місяць. Індекс містить інформацію більше ніж 8 млрд веб-сторінок на 186 мовах світу.

Google підтримує пошук документів у форматах PDF, RTF, PostScript, Microsoft Word, Microsoft Excel, Microsoft PowerPoint та ін.

Пошуковий робот Google має User Agent – Googlebot, який є основним роботом для сканування змісту сторінки для пошукового індексу. Крім нього існує ще декілька спеціалізованих роботів:

Googlebot-Mobile – робот, який індексує сайти для мобільних пристроїв;

Google Search Appliance (Google) gsa-crawler [5] – пошуковий робот нового апаратно-програмного комплексу Search Appliance;

Googlebot-Image – робот, котрий сканує сторінки для індексу картинок;

Mediapartners-Google – робот, який сканує зміст сторінки для визначення змісту AdSense;

Adsbot-Google – робот, що сканує контент для оцінки якості цільових сторінок AdWords.

Google використовує алгоритм розрахунку авторитетності

сторінки PageRank. PageRank є одним із допоміжних чинників при ранжируванні сайтів у результатах пошуку. PageRank не єдиний, але дуже важливий спосіб визначення розміщення сайту в результатах пошуку Google. Google використовує показник PageRank знайдених за запитом сторінок, щоб визначити порядок видачі цих сторінок користувачу в результатах пошуку.

Прихована на сторінках інформація, включно з мета-тегом Keywords не береться до уваги при ранжируванні сайтів.

Інтерфейс Google підтримує досить широке коло службових слів, що дають змогу обмежити ділянку пошуку окремими доменами, мовами, типами файлів і т. д.

17.2. Bing

Bing (бінг) – пошукова система, що належить компанії Microsoft. Має таку адресу: http://www.bing.com/. Почала працювати з 1 червня 2009 р. Незважаючи на третє місце в рейтингу пошукових машин світу (3,92 % станом на квітень 2011 р.), у нас маловідома й не використовується.

Раніше мала такі назви й адреси:

– MSN Search (http://search.msn.com/) з моменту появи восени 1998 р і до 11 вересня 2006 р.;

– Windows Live Search (http://search.live.com/) до 21 березня 2007 р.;

– Live Search (http://www.live.com/) до 1 червня 2009 р.

Пошукова система має багато унікальних особливостей, таких як: можливість перегляду результатів пошуку на одній веб-сторінці (замість того, щоб перегортати сторінки результатів пошуку), а також можливість динамічно корегувати обсяг інформації, що відображається для кожного результату пошуку (наприклад, тільки назва, коротка цитата або розширена цитата).

В американській версії Bing наявні деякі цікаві нововведення, що стосуються інтернет-пошуку, наприклад:

– теми оформлення стартової сторінки, які щодня змінюються, з інформаційними блоками;

– виведення уточнювальних варіантів пошукових запитів за окремими категоріями;

– відео з попереднім переглядом, що автоматично запускається;

– додаткові дані для кожного результату пошуку;

– вбудований сервіс для пошуку маршрутів (інші спеціальні пошукові сервіси з'являться незабаром);

– функції, які підвищують зручність у використанні під час пошуку інформації, зображень і відео.

Відразу ж після релізу пошукової системи з'явився жартівливий бекронім до слова "Bing" – "Bing Is Not Google".

17.3. AltaVista

AltaVista (www.altavista.com) – історично перша популярна і поширена пошукова система. Вона з'явилася у грудні 1995 р. і була підтримана найпотужнішим доступним на той час обчислювальним 64-бітовим сервером Альфа DEC. Це була найшвидша пошукова система, яка могла обробляти мільйони пошукових запитів за день. Назва AltaVista буквально перекладається як "погляд зверху".

Важливим нововведенням AltaVista було моделювання пошуку природньою мовою. Користувачі могли надрукувати фразу або питання й отримати інтелектуальну відповідь.

Раніше адреса пошукової системи була altavista.digital.com, і це збивало з пантелику багатьох користувачів Інтернету: набір в адресному рядку altavista.com приводив до потрапляння на сайт компанії AltaVista Technology. Як наслідок – популярність пошукової системи привела до величезного збільшення кількості відвідувачів на сайті AltaVista Technology і втрати потенційних користувачів пошуковою системою. У результаті, доменне ім'я altavista.com було викуплено за 3,35 млн дол. США у серпні 1998 р. (найбільша подібна операція на той час).

17.4. Yahoo!

Американську пошукову систему Yahoo! (www.yahoo.com) заснували аспіранти Стенфордского університету Девід Файло (англ. David Filo) і Джері Янг (англ. Jerry Yang), котрі у січні 1994 р. створили веб-сайт, який називався «Путівник Джері по Всесвітній Павутині». Путівником був каталог інших сайтів.

У квітні 1994 р. сайт перейменовано у Yahoo!. Існує дві версії походження назви: згідно з першою, слово було узяте з книги Джонатана Свіфта "Подорожі Гуллівера" (в українському перекладі, еху), де позначає расу грубих і тупих людиноподібних істот. Саме на цій версії наполягають

засновники компанії. Згідно з іншою, Yahoo! – акронім, утворений від фрази “Ще один ієрархічний офіціозний оракул” (англ. Yet Another Hierarchical Officious Oracle). URL сайту спочатку мав такий вигляд: http://akebono.stanford.edu/yahoo. Але є і третя версія походження назви. Цілком можливо, що Yahoo походить від японського Yahhoo, що в перекладі означає “Привіт” (неформальний варіант). Проте на той час Yahoo вже зареєстровано як торгову марку соусу для барбекю, тому до назви був доданий знак оклику. Компанія швидко розвивалася і стала корпорацією 2 березня 1995 р.

Yahoo! – друга за популярністю пошукова система світу (5,88 % станом на квітень 2011 р.). Вона надає багато сервісів, об’єднаних Інтернет-порталом Yahoo! Directory; портал містить популярний сервіс електронної пошти Yahoo! Mail, один із найстарших і найбільш популярних в Інтернеті. Не так давно була запущено нову версію поштового інтерфейсу, засновану на AJAX.

Наприкінці 2002 р. Yahoo! придбав інші пошукові системи Inktomi, а в 2003 р. Overture services Inc., AltaVista і AllTheWeb. 18 лютого 2004 р. Yahoo! припиняє використання пошукової технології Google і започатковує власну.

17.5. Yandex

Російська пошукова система Яндекс (www.yandex.ru) – найбільш популярна на сьогодні пошукова система на теренах колишнього Радянського Союзу. Ця пошукова система є восьмою серед найбільших пошукових сайтів світу за кількістю оброблених запитів (1,892 млрд – статистика за грудень 2009 р.) і другим найбільшим неангломовним пошуковим сервером після китайського Baidu. Пошукова система Yandex.ru була офіційно анонсована 23 вересня 1997 р. на виставці Softool, на початках розвивалася у межах компанії CompTek International. Як окрема компанія “Яндекс” утворилася 2000 р.

Основним і пріоритетним напрямом компанії є розробка пошукового механізму, але за роки роботи Яндекс став мульти-порталом. У 2009 р. в Яндексі доступні більше 30 сервісів. Найпопулярнішими є Яндекс.Картинки, Яндекс.Пошта, Яндекс.Погода, Яндекс.Новини та ін.

Назву “Яндекс” можна трактувати як скорочення від Yet

another indexer (ще один індексатор) або як “Мовний iNDEX”. Існує також такий варіант написання назви: “Яndex”.

17.6. Освітні програми Yandex

У 2004–2005 рр. і 2006–2007 рр. компанія фінансувала гранти на дослідження молодих учених із тематики інформаційного пошуку в межах конкурсу “Інтернет-математика”. На конкурс 2006–2007 рр. надійшло 156 заявок, із яких 31 відібраній заявці було виділено фінансування на загальну суму 5 млн рублів.

Конкурс “Інтернет-математика” проводився утретє у 2009 р., але в істотно зміненому форматі і з урізаним фінансуванням. Всі учасники вирішували одне загальне завдання, поставлене Яндексом: за виданими наборами даних, що складаються з оцінок релевантності деяких документів до пошукових запитів, потрібно було побудувати ранжирувальну формулу, за допомогою методів машинного навчання ранжируванню.

У 2007 р. була відкрита школа аналізу даних Яндексу, яка готує фахівців із прикладних галузей, пов’язаних з обробкою великих обсягів даних (зокрема, отриманих з Інтернету).

До 2009 р. Яндекс регулярно проводив змагання з пошуку в Інтернеті.

17.6.1. Пошук у Yandex

Пошукові сервіси Яндексу дають змогу шукати документи німецькою та французькою мовам з урахуванням морфології цих мов і близькості слів у реченні.

Крім веб-сторінок Яндекс індексує документи у форматах PDF (Adobe Acrobat), Rich Text Format (RTF), RSS (блоги та форуми) та бінарному форматі SWF (Macromedia Flash)

17.6.2. Мова пошукових запитів Yandex

Характерна особливість Яндексу – це можливість точного налаштування пошукового запиту, що реалізовано завдяки гнучкій мові запитів. Максимальний обсяг пошукового запиту може становити 40 слів.

17.6.3. Результати пошуку

Якщо за запитом знайдено дуже багато посилань, сторінка результатів пропонує обмежити діапазон пошуку за регіоном (тобто за діапазоном IP) або за датою. Якщо за вказаними

словами або словом нічого не знайдено, пропонується замінити їх на схожі; також можна виправити слова, набрані в неправильній розкладці клавіатури.

17.6.4. Якість пошуку

Час від часу алгоритми Яндексу, які відповідають за релевантність видачі, змінюються, що приводить до змін у результатах пошукових запитів. Востаннє така зміна відбулася у листопаді 2009 р., коли була викладена оновлена версія пошукової програми "Снєжинськ".

Яндекс підтримує низку різноманітних пошукових веб-сервісів, наприклад, пошук зображень, пошук у словниках, детальні карти міст, безкоштовний хостинг для персональних інтернет-сторінок, а також службу зберігання файлів; є навіть служба з генерації невеликих рефератів з основних наук.

17.7. Rambler

Російська пошукова система Рамблер (www.rambler.ru) – на сьогодні друга за популярністю пошукова система на теренах колишнього СРСР.

Пошукова система Рамблер розуміє і розрізняє слова російської, англійської та української мов. За замовчуванням пошук ведеться за всіма формами слова. Результати ранжируються за ступенем відповідності (релевантності) запиту і групуються на сайти; береться до уваги морфологія слів.

Цю систему створила група ентузіастів, що виникла у 1991 р. в підмосковному місті Пущино. Дмитро Крюков, Сергій Лисаков, Віктор Воронков, Володимир Самойлов, Юрій Єршов – саме ці технічно підковані люди стали засновниками Рамблера.

Компанія "Стек" на чолі з Сергієм Лисаковим зайнялася локальними мережами й підключенням до Інтернету. Нормальний, оперативний і ефективний обмін даними був потрібен для реалізації наукових цілей. "Стеківці" самі створили усередині міста мережу, підключили Пущино до Москви і через неї – до світового Інтернету (це був перший IP-канал у Росії, що виходить за межі Москви), приєдналися до мережі Курчатовського інституту атомної енергії. У 1992 р. компанія запустила власні ftp- і mail-сервери, через два роки – свій перший веб-сервер.

1996 рік знаковий для розвитку російського кіберпростору, адже

саме цього року Сергій Лисаков і Дмитро Крюков ухвалюють рішення розробити першу російську пошукову систему для Інтернету. Команда "Стека" зробила перспективну ставку на програму-пошукача. Це було важливе, відповідальне і на перший погляд необґрунтоване рішення, оскільки тоді російський сегмент Мережі був дуже малий – не більше 30–50 сайтів.

Програміст Дмитро Крюков почав із того, що придумав назву для проекту. Для цього він відкрив англо-російський словник, і погляд його впав на Rambler – Дмитру сподобалося значення слова "блукач, мандрівник, бродяга": роботи пошукової системи дійсно "блукають" по сайтах, збираючи інформацію цілодобово. Через декілька місяців, до осені 1996 р., був готовий остаточний варіант пошукача. 26 вересня зареєстровано домен rambler.ru, а 8 жовтня "Стек" активізував систему. У той момент у Рунеті уже існували дві-три пошукові системи, але вони не витримали перевірки часом і швидко зникли. А Рамблер розвивався, еволюціонував, запускав нові корисні сервіси.

Після необдуманої зміни інтерфейсу головної сторінки в 2007 р. пошукова система втратила свої позиції серед користувачів російськомовного Інтернету. Нині інтерфейс виправлений, але позиції втрачено – частина користувачів користується Яндексом.

17.8. Питання для самоперевірки

1. Які спільні ознаки мають пошукові машини загального користування?
2. Назвіть найбільш популярні пошукові системи світу.
3. Найпопулярніші вітчизняні пошукові машини.
4. Охарактеризуйте пошукову машину Google.
5. Особливості пошукової системи Bing.
6. Охарактеризуйте пошукову машину AltaVista.
7. Розкажіть про пошукову систему Yahoo!.
8. Історія створення й особливості пошукової системи Yandex.
9. Охарактеризуйте пошукову машину Rambler.

18. МЕТАПОШУКОВІ МАШИНИ

18.1. Nigma

Nigma (Нігма) (www.nigma.ru) – нова, але перспективна російська інтелектуальна метапошукова система, перша кластеризувальна пошукова система Рунету. Проект створений за

підтримки факультетів ВМК і психології МГУ, а також Стенфордського університету. Назва Nigma (один із видів павуків сімейства Nigma) була вибрана за асоціацією зі Всесвітньою павутиною.

На момент появи Nigma.ru у проекті брали участь троє людей, зараз у ньому працює більше 25 осіб.

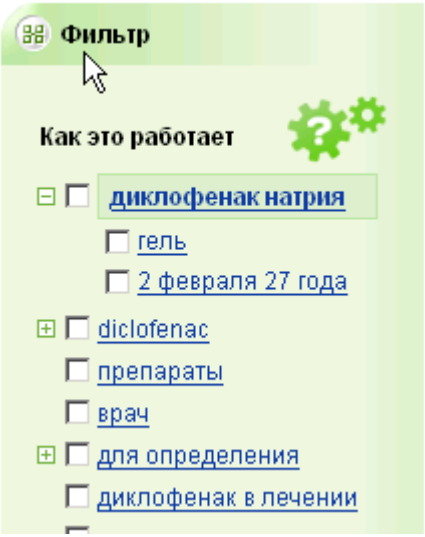
Нігма здійснює пошук як за своїм індексом, так і за індексами інших пошукових систем, причому є можливість обрання бази індексів для пошуку. Станом на 28 лютого 2009 р. в сумарному індексі всіх цих пошукових систем мітилося більше 7 160 000 000 російськомовних документів.

18.1.1. Кластеризація результатів пошуку

На основі введеного запиту Нігма формує список документів, розділених на декілька класів (кластерів). Користувач може уточнити, в якому класі продовжити пошук, поліпшивши так релевантність результатів пошуку. Користувач також може прибрати непотрібні йому класи сайтів, наприклад, документи з інтернет-магазинів (для них формується спеціальний кластер).

Список кластерів виводиться зліва від списку результатів пошуку. Користувач може управляти кластерами за допомогою спеціальних посилань під списком кластерів.

Приклад 52. *Кластеризація результатів пошуку в системі Nigma.ru:*



Нігма підтримує російську морфологію, використовуючи морфологічний модуль власної розробки. Раніше в Nigma.ru підтримка морфології була реалізована за допомогою відсилання в пошукові системи дублювальних запитів, у яких наведені поширені морфологічні форми слів-запитів. При цьому, на відміну від наявних реалізацій російської морфології для пошукових систем, пропонований алгоритм не скорочував, а збільшував кількість знайдених документів, оскільки морфологічно змінений запит об'єднується з початковим. Релевантність також збільшувалася, оскільки використовувалися спеціальні алгоритми об'єднання результатів.

Так, через Нігма, наприклад, можна було шукати документи в індексі Google з урахуванням російської морфології навіть тоді, коли Google не підтримував російську морфологію. Зараз потреба у цій технології зникла, оскільки всі провідні пошукові системи підтримують російську морфологію.

Підтримується орфографія на основі словника, складеного з реальної російської орфографії, яку автори веб-сайтів використовують на своїх сторінках, за винятком жаргону "падонкаф" (інтернет-спільнота, що використовує специфічний

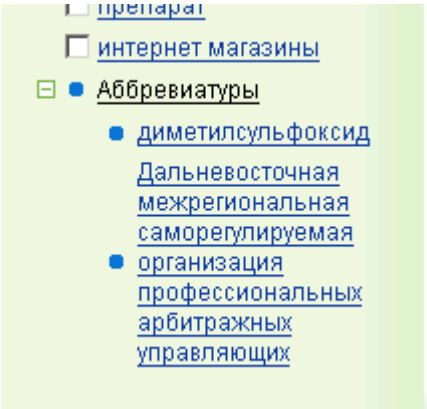
сленг, рясно приправлений ненормативною лексикою).

Система виправлення помилок Nigma.ru усуває друкарські помилки, пропонуючи користувачеві на вибір не один, а декілька варіантів виправлень, вона коректує також помилки, пов'язані з неправильним вибором розкладки клавіатури. Словник інтелектуальної пошукової системи розширений назвами відомих брендів, набирати які в рядку запиту користувач може навіть російською мовою, оскільки Нігма автоматично розширює пошук альтернативним написанням бренду.

18.1.2. Розшифрування скорочень

У Nigma.ru є сервіс розшифрування скорочень: скорочення вводять у рядок пошуку і паралельно з пошуком документів ведеться пошук розшифрування скорочень. Розшифрування, які отримали підтвердження у знайдених документах, потрапляють у список кластерів, а ті, що не отримали підтвердження в документах, потрапляють до спеціального псевдокластера "Абревіатури".

Приклад 53. *Результат розшифровки абревіатури "ДМСО" пошуковою машиною Nigma.ru:*



18.1.3. Специфічні сервіси й інструменти

Nigma.ru має низку специфічних сервісів, яких немає в інших пошукових системах. Найбільш корисними є математичні інструменти, конвертори величин та пошук за хімічними реакціями. Розглянемо ці сервіси більш детально.

18.1.3.1. Математичні інструменти

Система Нігма дає змогу не тільки проводити прості арифметичні перетворення, а й вирішувати математичні завдання різного ступеня складності. Також Нігма розпізнає більше тисячі фізичних, математичних констант і одиниць вимірювання, що дає змогу проводити операції з безліччю величин (зокрема вирішувати з ними рівняння) і отримувати відповідь у потрібних одиницях вимірювання. Крім рівнянь, система вирішує усі завдання, характерні для калькуляторів пошукових систем і конверторів валют. Нігма вміє проводити операції в дробах і розпізнає загальновживані синоніми валют. Так, наприклад, можна порахувати, скільки центів у доларі. За допомогою нового сервісу користувачі зможуть вирішувати різні математичні завдання (спрощувати вирази, розв'язувати лінійні і квадратні рівняння, системи рівнянь, рівняння з одиницями вимірювання, конвертувати валюти, обчислювати модуль числа, спрощувати тригонометричні вирази, скорочувати дроби та ін.), вводячи їх у рядок пошуку у вигляді строгого або нестрогого (звичайного) тексту.

Приклад 54. *Приклади математичних запитів до системи виразів Nigma.ru:*

- *8/13 + 2/3;*
- *cos2x + sin2x;*
- *2x – y = 4, 3y + x = 9;*
- *x + 1 = 0.*

На сьогодні розробники Нігма працюють над створенням додаткової можливості – виведення ходу рішення.

18.1.3.2. Хімічні реакції

Унікальною особливістю системи Нігма є пошук серед більш ніж 12 000 неорганічних реакцій. Речовини можна записувати як за допомогою назв (хлорид натрію, кам'яна сіль), так і у вигляді формул (NaCl). Для введеного набору з однієї або декількох речовин система спробує знайти реакції з їх участю.

Якщо користувач хоче знайти певну реакцію, він вводить речовини, розділяючи їх знаками "+", наприклад "гидроксид

натрия + HCl” (назви сполук на російській мові!). Система знайде всі реакції за участю NaOH і соляної кислоти. Розробники передбачили можливість вказати, з якого боку в реакції розміщені шукані речовини. Наприклад, якщо написати знак “дорівнює” після речовин (“2KOH + H2SO4 =”), то Нігма знайде тільки ті реакції, де гідроксид калію і сірчана кислота перебувають у початкових речовинах. Якщо користувач напише знак “дорівнює” перед реакцією: “= NaCl + H2S”, то Нігма знайде ті реакції, де в кінцевих продуктах є натрій хлорид і сірководень. Якщо користувач вкаже речовину і кінцевий продукт (“KOH = KCl”), то система відшукає всі реакції, у яких із гідроксиду калію одержують його хлорид. Окрім молекулярної формули для реакцій, що відбувається у розчинах, система видає іонну формулу, яка допоможе тим, хто вивчає хімію, краще зрозуміти суть хімічних процесів. У деяких випадках система може підказати користувачеві, чому потрібна реакція неможлива. Наприклад, на запит “K + H2O = K2O + H2” система дасть відповідь, що така реакція невідома, і, найімовірніше, відбуватися не може. Проте навіть хибні реакції урівнюються.

18.1.3.3. Пошук ланцюжків хімічних реакцій

Розробники Nigma.ru розширили функціональність пошуку за хімічними реакціями. Тепер можна шукати не одну реакцію, а відразу ланцюжок, наприклад: “NaCl = Na = NaH = NaOH = NaHSO3”. Система розіб’є ланцюжок на стадії і детально опише, як з однієї речовини отримати іншу. Для розділення стадій також можна використовувати стрілки, наприклад: “Fe -> FeS -> H2S -> S -> Na2S2O3 -> Na2S4O6”. У ланцюжках можуть бути і невідомі речовини, які можна позначати буквою “X”: “Ag -> X -> AgNO3 -> X -> Ag(NH3)2OH -> X -> Ag”, а також знаком питання і багатокрапкою: “Cu2O -> X -> CuSO4 -> ? -> CuCl2 -> … -> Cu2O”.

18.1.4. Автозаповнення рядка пошуку

По-перше, система пропонує варіанти на основі попередніх запитів користувачів Nigma.ru.

По-друге, напроти запропонованих прикладів автозаповнення

наведені найбільш релевантні сайти, пов'язані із запитом користувача. Тепер перейти на потрібний сайт можна навіть не дописуючи запит до кінця, а лише вибравши його зі списку і натиснувши клавішу "вправо" або просто натиснувши курсором на посилання. Якщо ж потрібний сайт стоїть на першому місці у підказці, то його можна не виділяти, а перейти на нього тільки за допомогою клавіші "вправо". Наприклад, якщо ввести дві букви: "по" і натиснути клавішу "вправо", то відкриється сайт gismeteo.

18.1.5. Допоміжні підказки

Вибираючи потрібний варіант в пошуковій підказці, користувач може побачити визначення понять і слів, які хоче знайти. База даних містить 340 928 визначень, інформацію взято з Вікіпедії; передбачається підключення і інших джерел інформації.

Одним із недоліків Nigma.ru є слабка сумісність із браузером "Internet Explorer" – на цьому сайті він часто зависає чи відключається. Рекомендуємо працювати з цим пошуковим сервером за допомогою браузера "Firefox".

18.2. Meta

Мета (www.meta.com.ua) – українська пошукова система, найвідоміший проект компанії ЗАТ "МЕТА" – розробника пошукових й інформаційних рішень. Сьогодні "МЕТА" – один із найбільш відвідуваних українцями сайтів і найбільший рекламний майданчик України.

"Мета.ua" – проект український, він створений і працює тільки для України. Пошукові технології компанії працюють у внутрішніх мережах Верховної Ради і Кабінету Міністрів України, на сайтах Національного банку України, фонду Разумкова.

Сервіси пошукової системи "МЕТА" можна поділити на три типи: пошукові, інформаційні та комунікаційні.

Із пошукових сервісів слід відзначити "Метановини". Зараз там зібрано новини більше ніж із 200 українських інтернет-джерел, близько 10 000 новин на день. Увесь цей обсяг у режимі реального часу індексується, групується за темами і стає доступним для пошуку.

18.2.1.1. Пошук рефератів

Майже єдиний сервіс у СНД, що дає змогу шукати не тільки за назвою та описом, а й за змістом тексту. У період сесій та іспитів студенти і школярі активно користуються цим сервісом.

З останніх пошукових проектів – інтерфейс до бази законодавства України, розроблений спільно з апаратом Верховної Ради. У базі більш як 80 000 різних юридичних документів. Автоматичний переклад запитів дає можливість формулювати запит російською або українською мовами.

Комунікаційні сервіси – форум, який став найбільшим українським неполітичним форумом. Поштовий сервіс розроблявся значно пізніше, ніж ті, що є зараз на ринку, тому в ньому вдалося обійти відомі недоліки, тому він зручний і функціональний. На сьогодні пошта – найпоширеніший сервіс на "Меті".

З останніх вдосконалень – "перевірка" правопису в запитах і добавлення нових форматів документів – doc, pdf, xls, ppt.

18.2.2. ПОРАДИ ДЛЯ ПОШУКУ

Пошукова система "МЕТА" пропонує низку сервісних можливостей, які дають змогу вести більш конкретний пошук інформації. Проте пошукова система – тільки інструмент, головний внесок у швидке отримання точних результатів робить користувач, коли формулює свій запит.

Нижче наведено перелік пошукових прийомів, які дають змогу ефективніше організувати цей процес і оперативно знайти потрібну інформацію.

18.2.2.1. Скільки слів використовувати у запиті?

За статистикою користувачі закордонних пошукових систем використовують у середньому 1,5 слова в запиті. Наші користувачі більш "багатослівні" – 2,5 слова на один запит.

Якщо потрібна загальна інформація, що якось стосується теми, достатньо одного слова. Напевно, серед декількох сотень документів, які видасть Мета, буде документ, який відповідає темі пошуку. Проте, де буде цей документ – у першій десятці результатів або десятій десятці – справа випадку.

18.2.3. Багатомовні запити

За статистикою велика частина запитів сформульована

російською мовою. При цьому пошукова база Мети містить документи російською, українською та англійською мовами.

Подібна багатомовність зумовлює свої особливості пошуку на Меті. Наприклад, для того, щоб отримати повний список сторінок, які стосуються освіти, слід крім слова "освіта" вказати також слова "образование" і "education". Якщо цікавить повнота пошуку, то це найбільш короткий спосіб, щоб отримати посилання на весь обсяг наявних документів.

Звичайно, багато з отриманих посилань міститиме однакову інформацію, яка просто представлена різними мовами. Проте, згідно з дослідженнями, значна кількість сторінок не повторюється, тобто містить інформацію, яка потрапляє в список результатів тільки при запиті певною мовою.

Однакове написання різних за сенсом слів (омонімія) при пошуку за ключовими словами може привести до появи у списку відповідей досить несподіваних результатів. Наприклад, до слова "лист" додатково до омонімії у російській мові: "лист каштана" і "лист бумаги" при пошуку на Меті додається ще значення "лист – письмо" українською. Тобто крім омонімії у російській та українській мовах окремо з'являється ще російсько-українська омонімія: приклад ружья – приклад перекладу, свято перемоги – свято верить, важкий стан – прокатний стан, Влада народу – позвать Влада і т. п.

Частково зняти подібну неоднозначність можна за допомогою оператора нормальної форми (поставити у запиті перед "підозрілим" словом знак оклику). Якщо ж використовувати пошук за фразою, омонімія мови майже не впливає на видачу результатів пошуку.

Пошукова система "МЕТА" дає змогу шукати у всьому українському Інтернеті, а також у Реєстрі українських сайтів.

Пошук відбувається, беручи до уваги російську та українську морфологію. Це означає, що незалежно від граматичної форми ключових слів, будуть отримані документи, які містять шукані слова у всіх формах.

Пошуковик має таку особливість: у багатослівних запитах система не ігнорує так звані "стоп-слова". Більшість пошукових систем при пошуку їх ігнорують, тобто при запиті "крем від засмаги" слово "від" буде проігноровано і серед результатів

будуть документи зі словосполученням “крем для засмаги”. “Мета” видасть документи, які точно збігаються із запитом.

Щоб покращити пошук, можна використовувати службові оператори. Порядок дії логічних операторів відзначається круглими дужками (). Фігурні дужки {...} дають змогу знаходити словосполучення, що є близькими до вказаного в них, тобто на відміну від попереднього оператора за запитом {погода в Криму} будуть знайдені документи, що містять словосполучення: “погода в Криму”, “погоді в Криму”, “погоди в Криму”, тобто граматична форма слів у такому разі не фіксується. Оператор [n, ...] використовують, якщо потрібно обмежити відстань між словами запиту. Наприклад, за запитом [5, мобільний телефон] будуть знайдені тільки ті документи, у яких слова “мобільний” і “телефон” розташовані у фрагменті тексту, що не перевищує п’яти слів. Передбачені також оператори, що дають змогу обмежити межі пошуку певним відрізком документа.

18.2.4. ПОШУК У РЕЄСТРІ УКРАЇНСЬКИХ САЙТІВ

Як і при повнотекстовому пошуку, пошук у Реєстрі ведеться із врахуванням російської, української й англійської морфології. За замовчуванням пошук ведеться у загальній базі даних; для пошуку в Реєстр слід поставити прапорець “шукати в Реєстрі”.

18.3. Питання для самоперевірки

1. Які спільні ознаки мають метапошукові системи?
2. Охарактеризуйте метапошукову системи Nigma.
3. Які можливості метапошукової системи Nigma у галузі хімії?
4. Охарактеризуйте метапошукову систему Meta.

19. ПОШУКОВІ МАШИНИ НАУКОВОЇ ІНФОРМАЦІЇ

19.1. Scirus

Scirus (http://www.scirus.com) – потужна пошукова машина, розроблена для пошуку лише наукової, академічної, технічної та медичної інформації. Обсяги наукової інформації й академічних знань, доступного завдяки Інтернету, фактично безмежні. Проте шлях до цієї інформації стає дедалі складнішим і забирає все більше часу. Тому компанія Elsevier розробила систему для пошуку винятково наукової інформації – Scirus.

Успіх Scirus полягає у тому, що:

– у пошуковій базі розміщені сайти з обмеженим доступом;

– у пошукову базу не внесені сайти без наукової інформації;

– вичерпний пошук у базах даних патентів, електронних архівів, депозитаріїв, препринтів та наукових статей;

– пошукова база містить інформацію про більше ніж 370 млн веб-сторінок;

– перевага надана рецензованим матеріалам;

– обробляється інформація, що міститься в нетекстових файлах типу PDF, PostScript та ін.;

– більш глибоке проникнення у структуру директорій веб-сайтів (на два рівні глибше, ніж у звичайних);

– завдяки більш повному аналізу індексованих документів видає більш точні результати.

Назва "Scirus" походить від назви провидця, що прийшов від богині Донони під час війни між грецькими містами Елейсисом та Афінами. Тут він був убитий, тому місце біля Елейсису було названо в його честь. Це ім'я для наукової пошукової системи вибрано через те, що вчені працюють із новими ідеями та розвитком, і вони є, на думку Elsevier, певною мірою провидцями.

З основних можливостей, що доступні за допомогою інтерфейсу Scirus, можна відзначити:

– вибір галузі знань: медицина, техніка, фізичні чи соціальні науки;

– уточнення пошуку конкретного автора, журналу чи статті;

– обмеження часового інтервалу за допомогою інтервалу дат;

– пошук наукових конференцій, рефератів і патентів;

– налаштування інтерфейсу та запис результатів пошуку;

– додаткові опції доступні через форму розширеного пошуку.

19.1.1. Технологія ранжирування результатів у Scirus

Зазвичай результати пошуку в Scirus ранжируються згідно з їх релевантністю. Порядок ранжирування можна змінити на ранжирування згідно з датою створення документа.

Релевантність документів у Scirus має дві складові частини:

а) частотність і розміщення слів у документі (статичний рейтинг);

б) кількість гіперзв'язків зі сторінкою: чим більше посилань,

тим більший рейтинг (динамічний рейтинг).

Загальний рейтинг є середнім арифметичним статичного та динамічного рейтингів. Для розрахунку релевантності не використовують метатеги.

До пошукової бази Scirus належать не всі сайти, а тільки ті, що відповідають певним вимогам: автоматично приєднуються сайти, які своєю адресою явно вказують на науковий зміст сайту (наприклад, www.newscientist.com), всі інші додають системні адміністратори або користувачі.

Як і більшість відомих пошукових систем, Scirus має свого робота – "павука", що сканує сайти, вказані в базі даних. Але, на відміну від інших роботів, "павук" Scirus досліджує посилання тільки тоді, якщо вони належать домену, який є в пошуковій базі сайтів. Завдяки значно меншому розміру пошукової бази, ніж у звичайних пошукових систем, робот-"павук" Scirus має змогу частіше відбирати нову інформацію в Інтернеті, тому його посилання на ресурси більш актуальні.

Якщо пройдена роботом сторінка потрапляє в пошукову базу Scirus, проводиться її автоматична предметна класифікація. На основі ретельного аналізу усіх слів сторінки відбувається віднесення її до однієї з 20-ти галузей науки. Документи, що перебувають на межі двох галузей, відносяться однозначно до обох. Класифікація за типом інформації проводиться залежно від типу документа.

19.1.2. Особливості пошукового інтерфейсу

19.1.2.1 Основна пошукова форма

В основній пошуковій формі доступні такі опції для уточнення пошуку:

– пошук тільки за точною фразою;

– показати в результатах пошуку тільки інформацію з журналів;

– показати в результатах пошуку інформацію з веб-сторінок.

На основі аналізу класифікаційних ознак перших 100 результатів пошуку складається список ключових слів, котрий зазначений праворуч сторони сторінки. При бажанні для подальшого уточнення пошуку можна вибрати одне з цих ключових слів або

форму розширеного пошуку.

19.1.2.2. Форма розширеного пошуку

У формі розширеного пошуку можна уточнити запит більш детально – шукати за певними частинами документів: у заголовку, серед авторів, у рефераті чи у повному тексті. Додаткові опції дають змогу:

– задати діапазон дат публікації;

– задати певні типи інформаційних джерел: усі джерела, реферати, книги, сайти компаній, конференції, патенти, препринти, домашні сторінки вчених;

– вибрати формат файлів: PDF, HTML або всі формати;

– уточнити джерела інформації: журнальні статті чи веб-сторінки, при цьому можна додатково вказати постачальника інформації;

– вибрати певні галузі знань з 20-ти можливих.

19.2. Питання для самоперевірки

1. Які особливості мають пошукові машини наукової інформації?
2. Як розраховується релевантність документів у Scirus?
3. Які можливості у пошуку наукової інформації має Scirus?

20. ПОШУКОВІ МАШИНИ БІБЛІОГРАФІЧНОЇ НАУКОВОЇ ІНФОРМАЦІЇ

20.1. SCOPUS

SCOPUS (http://www.scopus.com) – англомовний комерційний сайт, який підтримує базу даних наукової періодики. Проект був запущений 15 березня 2004 р. Це – бібліографічна і реферативна база даних та інструмент для відстеження цитованості статей, опублікованих у наукових виданнях. Індексує більше ніж 18 000 різних серіальних наукових видань із технічних, медичних та гуманітарних наук, більше ніж 5 000 видавництв та матеріали конференцій. Розробником і власником SCOPUS є видавнича корпорація Elsevier. База даних доступна за умови передплати через веб-інтерфейс. Пошуковий апарат SCOPUS інтегрований із пошуковою системою Scirus (для пошуку веб-сторінок) та патентною базою даних. Безкоштовним є тільки пошук за

авторами.

20.1.1. Тематико-типологічне покриття SCOPUS

Базу даних SCOPUS позиціонує видавнича корпорація Elsevier як найбільшу у світі універсальну реферативну базу з можливостями відстеження наукової цитованості публікацій. Згідно з оголошеною стратегією вона має стати найбільш повним та вичерпним ресурсом для пошуку наукової літератури. Станом на середину 2009 р. SCOPUS містив 38 млн записів наукових публікацій, включно 19 млн записів ресурсів, опублікованих після 1996 р. зі списками пристатейної бібліографії.

До класифікаційної системи SCOPUS входить 28 тематичних розділів:

– фізичні науки (32 %);
– хімічні технології;
– хімія;
– комп'ютерні науки;
– науки про Землю і планети;
– енергетика;
– виробництво;
– матеріалознавство;
– математика;
– фізика й астрономія;
– медичні науки (31 %);
– медицина та стоматологія;
– сестринська справа та медичні професії;
– фармакологія, токсикологія та фармацевтичні науки;
– ветеринарна справа та ветеринарна медицина;
– науки про життя (20 %);
– сільськогосподарські та біологічні науки;
– біохімія, генетика та молекулярна біологія;
– науки про навколишнє середовище;
– імунологія та мікробіологія;
– нейронауки;
– соціогуманітарні науки (17 %);
– мистецтвознавчі та гуманітарні науки;
– бізнес, менеджмент та бухгалтерський облік;
– теорії прийняття рішень;

– економіка, економетрика та фінанси;
– психологія;
– соціальні науки.

SCOPUS індексує наукові джерела, що видані різними мовами, за умови наявності у них англомовних версій рефератів. Географічне охоплення видавців за регіонами світу розподіляється так:

– Європа, Середній Схід та Африка (52 %);
– Північна Америка (36 %);
– Азійсько-Тихоокеанський регіон (9 %);
– Південна Америка (3 %).

20.1.2. Принципи відбору джерел для індексації у SCOPUS

Рішення про індексування нової назви видання базою даних SCOPUS приймаються за результатами розгляду запитів на внесення нових назв видань Консультативним комітетом SCOPUS з відбору змісту (CSAB). До цього комітету входять галузеві фахівці (близько 20 учених та 10 бібліотекарів), які репрезентують різні галузі знань та різні регіони світу. Запит на внесення нової назви видання до бази даних може подати будь-який учений за допомогою форми на веб-сайті ScopusInfo. Крім того, члени CSAB можуть самостійно визначати видання, які потрібно розглянути. Рішення про внесення нових назв видань до SCOPUS (та видалення назв видань, які не задовольняють вимоги SCOPUS) приймаються щорічно. Кінцевий термін подання науковцями запитів на внесення нової назви видання до SCOPUS для індексації у наступному році – 1 вересня кожного року. Видання, затверджені для внесення у базу даних, з'являються у SCOPUS на початку наступного року після прийняття запиту.

Для отримання інформації щодо видання, запропонованого для індексації у SCOPUS, експерти CSAB використовують дані про видання із запиту на внесення, матеріали англомовного сайту видання (якщо такий є) і дані каталогу періодичних видань Ulrich's Periodicals Directory.

20.1.3. Базові критерії оцінювання видання експертною радою SCOPUS

Критерії відбору видань для внесення до бази даних

SCOPUS такі:

– видання повинно мати англомовну назву та публікувати англомовні версії рефератів усіх наукових статей (оцінюється якість англомовних рефератів); повні тексти статей можуть публікуватися будь-якою мовою;

– періодичне видання повинно публікувати нові випуски з періодичністю не менш ніж один раз на рік;

– загальна якість видання повинна бути високою.

До критеріїв оцінки якості входять:

– авторитетність: включно з репутацією комерційного видавця або наукового товариства; різноманітність місць роботи авторів; міжнародний науковий авторитет провідних членів редколегії та різноманітність місць їхньої роботи (серед іншого, беруть до уваги цитованість членів редколегій та авторів журналів у виданнях, що вже індексує SCOPUS);

– популярність та доступність: разом з кількістю посилань на видання у базі даних SCOPUS; кількість установ, що передплачують видання; бази даних інформаційних агрегаторів, у яких вже індексується видання; кількість запитів на внесення видання до SCOPUS;

– діяльність видання повинна передбачати певну форму контролю за якістю публікацій (наприклад, наукове рецензування);

– видання повинно мати власний веб-сайт з англомовними версіями сторінок (оцінюється якість головної сторінки видання); наявність на веб-сайті повних текстів статей не є обов'язковою вимогою, але бажано для забезпечення можливостей безшовного переходу зі сторінок у SCOPUS до сторінок повних текстів статей ("View at Publisher").

Експертна рада CSAB обробляє усі запити на внесення нових назв видань до SCOPUS, але майже завжди автоматично відхиляє запити видань, які не мають:

– ISSN;

– стабільної регулярності виходу нових випусків;

– списків пристатейної бібліографії;

– англомовних рефератів до кожної статті;

– апарату рецензування;

– власного веб-сайту;

Здебільшого відхиляють запити на індексацію у базі даних

галузевих видань, що не відповідають критеріям відбору за типом документів, та запити на внесення журналів, котрі були додані до SCOPUS у 2004 р. (коли була запущена база даних), але у подальшому видалені з неї.

20.1.4. Наукометричний апарат SCOPUS

Наукові ресурси, опубліковані після 1996 р., індексуються у базі даних SCOPUS разом зі списками пристатейної бібліографії. Цитованість у базі даних підраховується шляхом автоматизованого аналізу змісту цих списків. Так, у SCOPUS підраховується кількість посилань на всі проіндексовані ресурси, але лише у виданнях опублікованих після 1996 р.

На відміну від бази даних Web of Knowledge Інституту наукової інформації США у SCOPUS не використовують поняття імпакт-факторів, натомість дуже широко застосовують індекс Хірша.

20.1.5. Профілі авторів

Для авторів, які опублікували більше однієї статті, у SCOPUS створюють індивідуальні облікові записи – профілі авторів з унікальними ідентифікаторами авторів (Author ID). Ці профілі надають таку інформацію: варіанти імені автора, перелік місць його роботи, кількість публікацій, роки публікаційної активності, галузі досліджень, посилання на основних співавторів, загальна кількість цитувань на публікації автора, загальна кількість джерел, на які посилається автор, індекс Хірша автора тощо. База даних надає користувачам можливість використання унікальних ідентифікаторів авторів для формування пошукових запитів та налаштування сповіщень (електронною поштою або через RSS) щодо змін у профілях авторів.

Можливості пошуку авторів та обмеженого перегляду їх профілів доступні без наявності передплати на базу даних SCOPUS засобами SCOPUS Author Preview.

20.1.6. Профілі установ

За аналогією з профілями авторів, для установ, співробітники яких опублікували більше однієї статті, у SCOPUS створюють профілі з унікальними ідентифікаторами установ (Scopus Affiliation Identifier). Ці профілі надають таку інформа-

цію: адреса установи, кількість авторів-співробітників установи, кількість публікацій співробітників, перелік основних назв видань, у яких публікуються співробітники установи та діаграма тематичного розподілу публікацій співробітників установи.

20.1.7. Профілі журналів

База даних SCOPUS надає широкі можливості отримання наукової метрики та автоматизованого аналізу видань. Інструмент Journal Analyzer дає змогу розширено аналізувати науковий рівень видань (включно з порівняльним аналізом декількох видань) за чотирма основними показниками:

– загальною кількістю статей, опублікованих у виданні протягом року;

– загальною кількістю посилань на видання в інших виданнях протягом року;

– трендом року (відношення кількості посилань на видання до кількості статей, опублікованих у ньому);

– відсотком статей, які не були процитовані.

20.1.8. Індексація українських наукових журналів у SCOPUS

Станом на 2010 р. у SCOPUS індексується зміст 35 українських журналів та одного збірника матеріалів конференції (15-та Міжнародна Кримська конференція “СВЧ – техника и телекоммуникационные технологии” (КриМиКо’2005), Севастополь, 12–16 вересня 2005 р.).

Із видань хімічного профілю у SCOPUS індексуються:

1. Сверхтвердые материалы. Інститут надтвердих матеріалів імені В. М. Бакуля НАН України, 1991–1998 рр.

2. Украинский химический журнал. Інститут загальної та неорганічної хімії імені В. І. Вернадського НАН України, 1972–1988, 1992–2005 рр.

3. Український біохімічний журнал. Інститут біохімії імені О. В. Палладіна НАН України, з 1965 р.

4. Фізико-хімічна механіка матеріалів. Фізико-механічний інститут імені Г. В. Карпенка НАН України, 1973, 1975–1976, 1991–1995, 2003–2005 рр.

5. Химия и технология воды. Інститут колоїдної хімії та хімії води імені А. В. Думанського НАН України, 1991–1998, 2000–2005 рр.

6. Biopolymers and cell. Інститут молекулярної біології і

генетики НАН України, з 2009 р.

Від початку роботи сервісу індексація більшості українських видань була припинена. Це, зокрема, серйозно вплинуло на показники України як країни-видавця. Аналогічна ситуація склалася з російськими виданнями та видавцями.

20.1.9. Використання даних SCOPUS у проектах оцінки наукової діяльності

База даних SCOPUS у багатьох країнах є одним із головних джерел отримання наукометричних даних для проведення оцінювальних досліджень на державному та корпоративному рівні.

Зокрема, у Російській Федерації відбулася перереєстрація усіх видань, внесених до "Переліку провідних рецензованих наукових журналів і видань, у яких повинні бути опубліковані основні наукові результати дисертацій на здобуття наукового ступеня доктора та кандидата наук", згідно з новими вимогами Вищої атестаційної комісії РФ. За новими вимогами ВАК РФ достатньою умовою для внесення наукового видання до "Переліку" є його індексація в одній з двох провідних світових баз даних відстеження цитованості: Web of Knowledge (Science Citation Index Expanded, Social Sciences Citation Index, Arts and Humanities Citation Index) або SCOPUS. Видання, які не індексуються у жодній з цих баз даних, для внесення до "Переліку" повинні відповідати низці додаткових критеріїв.

У січні 2010 р. Санкт-Петербурзький державний університет ввів у дію наказ "Щодо врахування цитованості наукових та учбово-методичних робіт при заміщенні посад науково-педагогічних працівників у СПбДУ". Згідно з ним претендентам на заміщення посад викладачів потрібно надати дані щодо цитованості робіт, опублікованих ними за останні п'ять років, у наукометричних базах даних Web of Knowledge, Scopus та РИНЦ.

Університет Гранади (Іспанія) на основі даних SCOPUS підтримує проект SCImago Journal & Country Rank (SJR), суть якого полягає у публікації науково-аналітичних звітів про журнали та країни, а також рейтинги публікаційної активності та статистики цитованості журналів і країн світу. У рейтингу SJR станом на 2007 р. в цілому у всіх галузях знань за кількістю публікацій Україна посідає 33 місце, а за кількістю цитувань –

40-ве. Індекс Хірша України за даними SJR 2007 становить 94 (тобто українські вчені опублікували 94 статті, кожна з яких була процитована принаймні 94 рази) – 43-тє місце серед країн світу. За кількістю публікацій найвищий рейтинг Україна має у галузі механіки матеріалів (9-те місце серед країн світу), найнижчий – у галузі стоматології (110-те місце). На основі даних SCImago Journal & Country Rank розпочато проект створення Атласу науки, у якому має бути графічно реперзентована світова структура наукових досліджень. Нині ведеться робота над Атласами науки Іспанії, Португалії та восьми країн Південної Америки.

Дані SCOPUS використовують у рейтингу провідних університетів світу Times Higher Education Supplement: World University Rankings (QS TopUniversities). Українських університетів у цьому рейтингу немає.

Компанія Academic Analytics використовує дані SCOPUS для підрахунку Faculty Scholarly Productivity Index (метричного показника оцінки якості наукової діяльності університетів США).

Проекти щодо використання наукометричної бази даних SCOPUS для оцінки наукового потенціалу держави в цілому та окремих суб'єктів наукової діяльності в Україні на рівні вищих органів державної влади почали з'являтися у 2009 р. Протягом року Міністерство освіти і науки України та Національна академія наук України провели низку консультацій із вищим керівництвом компанії Elsevier щодо серйозного збільшення номенклатури українських видань, що індексуються у SCOPUS (до показника 100–200 журналів; на рівні 40–80 % аналогічного показника Польщі). Стратегічні домовленості з цього питання були досягнуті. Органи державної влади та профільні відомства розпочали процеси роботи з видавцями щодо цільової підготовки наукових фахових видань України для внесення у SCOPUS і паралельно процеси запровадження у практику власної діяльності методів кількісної та напівкількісної оцінки наукового потенціалу (наукометрії) на основі даних Scopus.

10 червня 2009 р. Президія НАН України ухвалила рішення щодо розгортання системи постійного моніторингу наукового потенціалу суб'єктів наукової діяльності України за показниками бази даних SCOPUS. Реалізація цього завдання

була покладена на Національну бібліотеку України імені В. І. Вернадського.

26 жовтня 2009 р. Комітет із Державних премій України в галузі науки і техніки затвердив нові редакції "Інструкцій про порядок висунення, оформлення та представлення робіт на здобуття Державних премій України в галузі науки і техніки та на здобуття щорічних премій Президента України для молодих вчених". Згідно з новими редакціями цих інструкцій претенденти на здобуття премій у анотаціях робіт зобов'язані вказувати, зокрема, загальну кількість статей, опублікованих у міжнародних журналах, що містяться в базі даних SCOPUS, загальний ідентифікатор SJR (SCImago Journal Rank) та загальний індекс цитування робіт претендентів.

24 грудня 2009 р. рішенням колегії Міністерства освіти і науки України показник "Кількість публікацій у наукометричній міжнародній базі даних SCOPUS" був ухвалений як один із показників оцінки результативності наукової та науково-технічної діяльності вищих навчальних закладів.

20.2. Питання для самоперевірки

1. Які особливості пошукових машин бібліографічної інформації?
2. Тематико-типологічне покриття SCOPUS.
3. Принципи відбору джерел для індексації у SCOPUS.
4. Базові критерії оцінювання видання експертною радою SCOPUS.
5. Наукометричний апарат SCOPUS.
6. Що таке профілі у SCOPUS?
7. Які є види профілів у SCOPUS?
8. Які українські журнали індексуються у SCOPUS?

21. ІНТЕРНЕТ-ЕНЦИКЛОПЕДІЇ

21.1. Вікіпедія

Ларрі Сенгер, колишній головний редактор Nupedia (онлайн-енциклопедія – попередник Вікіпедії), запропонував ідею Вікіпедії і частину з її основоположних правил. Вікіпедія – онлайн-енциклопедія з можливістю вільного редагування.

Офіційно відкрита 15 січня 2001 р. Спочатку вона була створена як доповнення до вільної енциклопедії Nupedia, яку писали експерти, для створення додаткового джерела чорнових

статей та ідей. Вікіпедія швидко зайняла позиції Nupedia і стала великим міжнародним проектом, що концентрує навколо себе безліч додаткових проектів. У 2008 р. Вікіпедія містила більше 10 млн статей, написаних мільйонами її учасників. Вона є одним із найпопулярніших веб-сайтів і широко використовується як довідник.

21.1.1. Огляд історії створення Вікіпедії

Концепція збору всіх усесвітніх знань в одному місці походить від стародавньої Александрійської бібліотеки і Пергамону. Ідея використовувати автоматизовані системи поза друкарською машиною, щоб створити кориснішу енциклопедію, висловлена у статті бібліотекаря Чарльза Аммі Каттера "The Buffalo Public Library in 1983" (Library Journal, 1883), книзі Пола Відліта "Traité de documentation" (1934) (Відліт також заснував установу Mundaneum, 1910); збірці есе Г. Г. Уеллса "World Brain" (1937) й у футуристичному мікрофільмі Ванневара Буша, заснованого на Memex з As We May Think (1945). Іншою віхою був Project Xanadu Теда Нельсона 1973 р.

Із розвитком Інтернету багато вчених намагалися розвинути проекти інтернет-енциклопедії. Фахівець вільного програмного забезпечення Річард Столлман визначив повноцінність "Вільної універсальної енциклопедії" у 1999 р. Він описав "цілі, які повинна ставити вільна енциклопедія, які свободи вона повинна дати суспільству і як можна почати її розробляти". Фонд вільного програмного забезпечення заохочує людей "відвідувати сайт і вкладати у нього свій внесок". Одним із маловідомих попередників була Interpedia, яку Роберт Макгенрі концептуально пов'язував з Вікіпедією.

21.1.2. Формулювання концепції

Вікіпедія була спочатку задумана як проект-наступник Нупедії – проекту, заснованого Джиммі Уейлсом, щоб провести вільну енциклопедію. Нупедія була заснована на участі висококомпетентних учасників і складного багатокрокового процесу рівноправної рецензії. Незважаючи на список зацікавлених розробників і присутність постійного головного редактора, Ларрі Сенгера – дипломованого філософа, створення відбувалося дуже повільно: протягом першого року було створено лише 12 статей.

Уейлс і Сенгер обговорювали різні способи пришвидшення процесу складання змісту. Ідея заснованого на Вікі доповнення з'явилася після розмови між Ларрі Сенгером і Беном Ковітзом. Бен Ковітз, програміст і постійний клієнт Ward Cunningham's wiki (the WikiWikiWeb), представив Сенгеру wiki 2 січня 2001 р. У жовтні 2001 р. Уейлс заявив, що "у Ларрі була ідея використовувати програмне забезпечення Wiki". (Пізніше, у грудні 2005 р. він стверджував, що Джеремі Розенфелд, службовець Bomis, представив йому цю концепцію.) Сенгер вважав, що Вікі буде хорошою платформою, і запропонував учасникам Нупедії встановити Вікі, засноване на UseModWiki (тоді v. 0.90), як наступника Nupedia.

Такі зміни спричинили значний опір з боку редакторів і рецензентів Нупедії. Сенгер запропонував дати новому проектові його власну назву – "Вікіпедія" і 15 січня він був запущений під цим ім'ям на власному домені wikipedia.com.

Сервер (розміщений у Сан-Дієго), використовуваний для цих проектів, був безкоштовно наданий Bomis. Багато нинішніх і колишніх службовців Bomis зробили свій внесок у розширення обсягу енциклопедії, особливо Тім Шелл – співзасновник і на сьогодні головний адміністратор Bomis, і Джейсон Річі – програміст.

Традиційно вважають, що першими змінами, зробленими у Вікіпедії, є пробні правки Уейлса. Проте найдавніша стаття з тих, що існують і нині, – це стаття про UUU, створена 16 січня 2001 р., о 21:08 UTC.

У проекту з'явилося багато нових прихильників після триразової згадки на веб-сайті Slashdot і двох незначних згадок у березні 2001 р. Крім цього істотний потік відвідувачів з'явився і з інших сайтів, серед яких варто відзначити Google, котрий щодня додавав сайту сотні нових відвідувачів. У пресі сайт вперше був висвітлений у "Нью-Йорк Таймс" 20 вересня 2001 р.

Кількість статей у проекті перевищила 1000 12 лютого 2001 р., а 7 вересня – 10 000. Протягом першого року існування Вікіпедії було створено більше 20 000 енциклопедичних статей – в середньому більше 1500 статей на місяць. 30 серпня 2002 р. їх кількість досягала 40 000. Зростання більш-менш стійко відбувалося від запуску проекту, за винятком декількох уповільнень, викликаних збоями у програмному забезпеченні й

апаратурі.

21.1.3. Інтернаціоналізація

Вікіпедія вже на початкових етапах свого розвитку почала розширюватися інтернаціонально, створюючи нові мовні розділи. Перший домен, зарезервований для неанглійської Вікіпедії, був deutsche.wikipedia.com (16 березня 2001 р.); протягом приблизно двох місяців він був єдиним, що містив статті неанглійською мовою. Перші відомості про французьку Вікіпедію з'явилися 23 березня 2001 р., а починаючи з травня 2001 р. створені розділи китайською, нідерландською, есперанто, івритом, італійською, японською, португальською, російською, іспанською і шведською мовами. Дещо пізніше приєднали арабську й угорську мови. У вересні 2001 р. було висунуто обов'язкову умову до багатомовності Вікіпедії, включно з відходом від усіх головних мов і встановленням основних стандартів, перекладу основних сторінок для нових Вікі. У кінці 2001 р. було оголошено про африкансько-, норвезько-, і сербськомовні підрозділи.

У січні 2002 р. 90 % усіх статей Вікіпедії були англійською мовою. До січня 2004 р. їх частка зменшилася до 50 %, й інтернаціоналізація далі зростала. У 2007 р. близько 75 % усіх статей містилися у межах неанглійських розділів Вікіпедії.

21.1.4. Організація

Із початку створення і донині триває розширення проекту Вікіпедії на різних рівнях, зростає кількість статей у наявних проектах, з'являються нові проекти, відбувається удосконалення програмного забезпечення. За час розвитку з'явилася можливість безпосереднього введення математичних формул мовою TEX. Для написання статей почали використовувати боти (програми, схожі на пошукові павуки). ISBN у статтях тепер дають посилання на сторінку Special:Booksources, яке приводить їх зміст із джерел сторінки Wikipedia:Book.

Якщо Вікіпедія розвиватиметься такими ж темпами, як зараз, можна очікувати, що до 2020 р. вона посідатиме 4-те місце за популярністю в Інтернеті після новин, чатів й офіційних сайтів фірм і підприємств.

Були випадки, коли Вікіпедію блокувала національна влада. І нині вони пов'язані з КНР, Іраном, Тунісом, Узбекистаном і Сирією.

Китайська Вікіпедія блокувалася вже неодноразово. Уперше доступ користувачів з КНР до серверів Вікіпедії був заблокований 2–21 червня 2004 р. це пов'язано з 15-ю річницею подій на площі Тяньаньмень 4 червня 1989 р. Удруге такий випадок стався 23–27 вересня 2004 р. Проте це обмеження не мало всеосяжного характеру і торкнулося не всіх користувачів Китаю. Причини цього блокування досі невідомі. Третє блокування тривало з 19 жовтня 2005 р по 10 листопада 2006 р.

Доступ до Вікіпедії для користувачів з Ірану був закритий 3 грудня 2006 р., одночасно з іншими популярними сайтами, наприклад, YouTube і IMDB. Вважається, що це блокування пов'язане з розгорненою в країні кампанією щодо боротьби із впливом західної культури.

Веб-сайт Вікіпедія був заблокований на декілька днів (23 – 27 листопада 2006 р.) в Тунісі.

Доступ до арабської Вікіпедії був закритий з 30 квітня 2008 р. і досі не знятий.

21.2. Питання для самоперевірки

1. Що таке Вікіпедія?
2. На яких засадах існує Вікіпедія?
3. Яка користь Вікіпедії під час пошуку наукової інформації?

22. ВЕБ-АРХІВИ

Інтернет-архів (www.archive.org, Internet Archive) – неприбуткова організація, розміщена у Сан-Франциско, штат Каліфорнія, для збереження культурного надбання. Архів підтримує електронну бібліотеку, архів мережевих і мультимедійних ресурсів, програмного забезпечення, фільмів, книг і звукозаписів. Архів "Wayback Machine" містить копії сторінок веб-сайтів з різних проміжків часу (зазвичай кожні два місяці). Щоб гарантувати стабільність архіву, веб-дзеркало Інтернет-архіву також зберігається в Новій Олександрійській бібліотеці в Єгипті. Інтернет-архів безкоштовний, з 2007 р. входить до складу Американської асоціації бібліотек й офіційно визнаний у штаті Каліфорнія як бібліотека.

Некомерційну організацію Internet Archive заснував Брюстер Калі (Brewster Kahle) у 1996 р. для створення загальнодоступної електронної бібліотеки. Частиною цієї роботи є Wayback Machine – колекція архівних копій веб-сторінок.

Wayback Machine – архів, що складається з копій сторінок веб-сайтів, які взяті з Alexa Internet. Цей архів дає змогу користувачам бачити версії веб-сторінок у тому вигляді, у якому вони були подані у різні проміжки часу.

Архів Wayback Machine стає доступним не одразу: період з моменту копіювання та архівації веб-сторінки до моменту, коли вона стає доступною у Wayback Machine, може становити близько шести-дванадцяти місяців. Як альтернатива, користувачі, які хочуть заархівувати матеріал і негайно його процитувати, можуть скористатися Archive-it.org. Станом на 2006 р. Wayback Machine містила майже два петабайти даних, збільшуючись на 20 терабайтів щомісяця.

22.1. Мультимедійні ресурси

Як додаток до мережевих архівів, Інтернет-архів підтримує обширну колекцію цифрових мультимедійних ресурсів, які або є суспільним надбанням, або розповсюджуються за Creative Commons License чи еквівалентною ліцензією, яка дозволяє вільне використання та розповсюдження.

Сервери веб-дзеркала Архіву в Олександрії – Open Library (Вільна бібліотека) – нараховує більше мільйона оцифрованих книжок, які є суспільним надбанням у США.

22.1.1. BookServer

Некомерційний проект Internet Archive восени 2009 р. оголосив про запуск BookServer – універсальної і відкритої системи дистрибуції електронних книг. BookServer повинен стати архітектурою, яка об’єднає різні формати і пристрої для читання електронних текстів. Система забезпечить каталогізацію усіх книг, наявних у магазинах, бібліотеках або у відкритому доступі. Пошукові системи зможуть проіндексувати цей каталог, і читачам стане значно простіше шукати потрібну інформацію.

На час оголошення у 20-ти центрах сканування Internet Archive, розміщених у п’яти країнах світу, тривала робота з оцифровування книг, призначених для відкритого доступу, і цей

каталог налічував більше 1,6 млн найменувань. Але приблизно половина з них – книги, відсканованої Google у межах проекту Book Search і завантажені згодом користувачами на сайт Internet Archive.

Попередня версія системи BookServer уже працює на сайті власне Internet Archive і в проекті з дистрибуції електронних книг Feedbooks. Видавництво O'Reilly Media, яке надавало допомогу в реалізації проекту, планує внести до системи усі свої книги. З BookServer співпрацюють також компанія Adobe, університет Торонто і проект "One Laptop Per Child".

22.2. Питання для самоперевірки

1. Охарактеризуйте проект Internet Archive.
2. Яку інформацію можна знайти у веб-архіві?

23. ХІМІЧНІ КАТАЛОГИ

23.1. Wiley Interscience

Wiley InterScience (http://www3.interscience.wiley.com) – загальнонауковий онлайн-каталог видавництва "John Wiley & Sons, Inc" (Wiley), що забезпечує доступ до близько 3 млн статей із 1500 журналів та більш ніж 7000 електронних книг, один із розділів якого присвячений хімії. Підтримується доступ також до баз даних з хімії.

24.ХІМІЧНІ ПОРТАЛИ

24.1. ChemWeb

ChemWeb (ChemWeb.com) забезпечує хімікам доступ до інформації, якої вони потребують. Сюди входить доступ до рефератів, статей, журналів, книг, конференцій, форумів з хімії, сервера новин Alchemist й низки баз даних. Пошукові сервіси містять пошук робіт, книг, інформації про конференції, онлайн-магазини.

Сервіс новин Alchemist запущений у липні 1997 р. і забезпечує всіх учасників ChemWeb новинами в галузі хімії та пов'язаних із нею дисциплін.

Для доступу до сервісів ChemWeb потрібно зареєструватися на сайті, після чого можна вибрати мову інтерфейсу; частина елементів

буде українською мовою (якщо вибрано українську мову).

БАЗИ ДАНИХ ІНФОРМАЦІЇ З ХІМІЇ

PubChem

PubChem (http://pubchem.ncbi.nlm.nih.gov) – база даних хімічних молекул. Цю систему підтримує Національний центр біотехнологічної інформації США NCBI (англ. National Center for Biotechnology Information), котрий є частиною Національної медичної бібліотеки США і Національного інституту здоров'я США (NIH). Довільний доступ можливий за допомогою веб-інтерфейсу PubChem.

Метою PubChem є забезпечення інформацією про біологічну активність невеликих молекул; його бази даних містять інформацію про речовини, структури сполук, дані про біологічну активність. Обсяг бази PubChem досить великий, містить більш ніж 62 млн записів про окремі речовини, більше ніж 26 млн унікальних структур та більш ніж 2000 біоматриць.

Пошукова форма PubChem містить значну кількість опцій, дає змогу використовувати для уточнення багато критеріїв. Проте інформація, одержана як результат пошуку, погляду хіміка, досить примітивна: молекулярна маса, молекулярна формула, двовимірна та тривимірна структури молекули.

Позитивним аспектом PubChem є наявність гіперпосилань на більшість елементів з результату пошуку, можливість шукати подібні за структурою сполуки, при цьому для представлення ступеня схожості використовують кластерний аналіз і будують ієрархічне дерево, на якому можна знайти схожі за будовою сполуки. Для кожної знайденої сполуки поданий весь набір хімічних ідентифікаторів сполук: SMILES, CAS#, InChi, InChiKey.

Як недоліки можна відзначити те, що часто одна і та ж сполука представлена в базі даних багато разів, в деяких випадках внесені некоректні структурні формули:

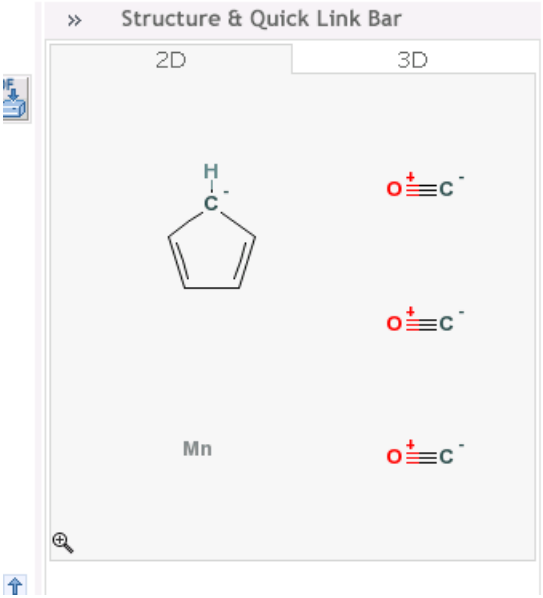


Рис. 19. *Приклад некоректної структури цимантрену в PubChem*

ChemSpider

ChemSpider (http://www.chemspider.com) – система для пошуку інформації з хімії. Мета сайту – зібрати в одному місці інформацію про хімічні сполуки, їх структури і супутню інформацію та зробити її доступною для усіх користувачів. Цей сайт був запущений у березні 2007 р. й підтримується Королівським хімічним товариством Великобританії, права на нього належать RSC Worldwide Ltd.

Засновники і власники ChemSpider заявляють, що не несуть ніякої відповідальності за представлену на сайті інформацію з хімії та за наслідки її використання.

Для коректної роботи із сайтом потрібні браузери Internet Explorer, починаючи з сьомої версії або Firefox.

За допомогою пошукової форми ChemSpider можна проводити пошук за ключовими словами або за фрагментами хімічних структур. База даних містить інформацію про структури і властивості мільйонів різних сполук.

Для знайдених сполук обов'язково вказані систематична назва, тривіальна назва, хімічні ідентифікатори SMILES, InChI, InChIKey, CAS#. Наводяться передбачені фізико-хімічні властивості, котрі розраховані за допомогою програмного забезпечення ACD/Lab.

Оскільки веб-робот сайту аналізує бази даних, які мають непрямий стосунок до хімії, а інформацію на сайті може розмістити будь-хто з користувачів без попереднього рецензування, але часто вона має відверто некоректний характер. Робота сайту в цьому аспекті дещо схожа з роботою Вікіпедії, де будь-хто може розмістити будь-що. Хоча тут діє механізм, подібний до Вікіпедії, – будь-який користувач може вказати на некоректну інформацію, тоді вона буде видалена чи скоректована.

Перевага сайту полягає у тому, що на будь-яку фактичну інформацію подані посилання або гіперпосилання на джерело, зазвичай це певні журнальні публікації.

Незважаючи на певну кількість некоректних даних, цей сайт може бути корисним у процесі попереднього пошуку фактичної інформації та визначення першоджерел.

NIST

NIST (англ. The National Institute of Standards and Technology (NIST)) – Національний інститут стандартів і технологій (США). Місія Інституту – "просувати" інноваційну й індустріальну конкурентоспроможність США завдяки розвитку стандартизації і технологій для підвищення економічної безпеки і поліпшення якості життя.

У NIST міститься низка баз даних фактичної наукової інформації, до деяких із них є вільний доступ. До більшості доступ платний, проте є демо-версії, які не містять найновішої інформації.

В Інтернеті доступ до баз даних надається через веб-портал http://webbook.nist.gov. Інформації з хімії присвячено розділ "NIST Chemistry WebBook" (http://webbook.nist.gov/chemistry).

The NIST Chemistry WebBook

The Chemistry WebBook (Веб-книга NIST з хімії) забезпечує

доступ до даних, які зібрані та поширюються у NIST у межах програми “Standard Reference Data Program”.

“The NIST Chemistry WebBook” містить термохімічні дані для більш ніж 7000 органічних та неорганічних речовин, термохімічні дані для більше ніж 8000 реакцій, ІЧ-спектри для більше ніж 16 000 сполук, мас-спектри для понад 15 000 сполук, УФ-спектри більш ніж 1600 сполук, дані про газову хроматографію для понад 27 000 сполук, електронні та коливні спектри більш ніж 5000 сполук, константи двоатомних молекул (спектроскопічні дані) більше 600 сполук, енергетичні характеристики більше ніж 16 000 сполук, термофізичні властивості для 74 рідин.

У базі можна шукати дані про певні сполуки, ґрунтуючись на назві, хімічній формулі, CAS registry number, молекулярній масі, хімічній структурі, енергетичних або спектральних властивостях молекул.

NIST попереджає, що може в будь-який момент закрити доступ до баз даних, але поки цей доступ є. За його словами, інформація у базах даних перевірена, хоча й ніякої відповідальності за можливі помилки NIST не несе.

Оскільки база даних NIST Chemistry WebBook є однією з небагатьох, котра надає вільний доступ до фактичного матеріалу, то розглянемо її особливості більш детально.

A Guide to the NIST Chemistry WebBook

Пошук у базах даних NIST можна проводити двома способами: прямий пошук за хімічною сполукою і непрямий пошук, що ґрунтується на відомих властивостях сполук.

Залежно від того, скільки результатів знайдено під час пошуку, видається або кінцева сторінка (у разі одного результату), або список знайдених сторінок (у разі декількох результатів).

Зміст цієї сторінки залежить від кількості варіантів. Обов’язково зазначають назви й хімічні формули знайдених сполук; якщо результатів небагато, будуть вказані також структурні формули. Зі сторінки результатів пошуку за гіперпосиланнями можна перейти до сторінок, що містять більш

детальну інформацію.

Безпосередній пошук

NIST Chemistry WebBook підтримує декілька видів безпосереднього пошуку хімічних сполук, кожен із яких має власну пошукову сторінку. Нині доступні такі види пошуку: за хімічною формулою; за назвою; за номером CAS; за енергією іонізації; за спорідненістю до електрона; за спорідненістю до протона; за кислотністю; за енергією утворення; за коливальною енергією; за енергією електронних рівнів; за структурними фрагментами; на основі структурних файлів; на основі структурних класів; за молярною масою; пошук реакцій; авторський пошук.

Більшість видів пошуку мають спільні особливості – це можливість вибирати одиниці вимірювання, формат знайдених результатів.

Пошук фізичних властивостей можна обмежити класами сполук, наприклад, вуглеводнями чи амінами.

Під час безпосереднього пошуку можна вибрати одиниці, у яких будуть представлені результати у системі СІ чи заснованої на калоріях.

У більшості типів пошуку потрібно вказати, які дані слід вивести на сторінці результатів. Безпосередній пошук підтримує такі опції:

– термохімічні дані в газовій фазі;
– термохімічні дані у твердому стані;
– дані про фазові переходи;
– термохімічні дані реакцій;
– енергію іонів у газовій фазі;
– дані про кластеризацію іонів;
– ІЧ-спектри в газовій фазі;
– мас-спектри;
– УФ-спектри;
– коливальні та електронні спектри;
– константи двоатомних молекул;
– константи Генрі.

При успішному пошукові будуть знайдені дані відповідно до вибраних опцій. Якщо нічого не відзначено, буде видана тільки загальна інформація про сполуку.

При виділенні тільки певних категорій для пошуку потрібно використати логічне “I”, і будуть видані результати за тими сполуками, що містять весь набір вибраних даних.

Пошук за формулою

При пошуку за формулою будуть знайдені усі сполуки, що містять вказаний у пошуковій формі фрагмент. Формула може бути записана у будь-якому порядку. Доступні символи – великі і малі букви, цифри, знак питання, лапки й дужки. Усі спроби розписати структуру будуть проігноровані – пошук вестиметься тільки за брутто-формулою.

Приклад 55. *Спрощення пошукового терміна в NIST Chemistry WebBook:*

За запитом “(CH3)(CH2)5(CH3)” будуть знайдені сполуки, які мають формулу C7H16, а не тільки н-гептан.

Набираючи формули, потрібно дотримуватися встановлених символів елементів, тому що можна одержати неочікувані результати.

Під час пошуку можна використовувати додаткові знаки, які вказують на кількість атомів, при цьому у формулі не можна використовувати дужки. Як додаткові знаки можна використовувати знак питання чи зірочку.

Знак питання (?) означає один чи більше атомів елемента, що передує знакові. Наприклад, CH? – будуть знайдені усі сполуки з одним атомом вуглецю й одним чи більше атомів водню.

Зірочка (*) означає будь-яку кількість атомів елемента, що передує знакові. Його дія така ж, як і знака питання, може означати й відсутність вказаного елемента.

Приклад 56. *Використання додаткових знаків у пошукових запитах до NIST Chemistry WebBook:*

*“C2C*H4” – будуть знайдені сполуки, що містять чотири атоми водню і два або більше атоми вуглецю.*

У цьому виді пошуку можна його уточнити, відзначивши потрібні опції:

– точну відповідність зазначеним ізотопам;

– дозволити елементи, не вказані у формулі. Ця опція значно збільшує кількість результатів. Небажані елементи можна замінити, вказавши нуль після символа елемента, наприклад, “CH2N0S0”;

– дозволити більше атомів кожного зі вказаних елементів;

– усунути з пошуку іони.

Пошук за хімічною назвою

При пошуку за хімічною назвою (Chemical Name Search) ігнорується регістр букв, а також пробіли в запиті: “hydrogenfluoride” і “Hydrogen Fluoride” будуть означати одну і ту ж сполуку HF. У базі даних, крім хімічних назв, є також комерційні назви й акроніми.

У запитах можна використовувати додатковий знак зірочки (*), яка замінює будь-який текст.

Якщо за словом запиту нічого не знайдено, то пошук автоматично проводиться в базі даних формул, а потім – у базі CAS registry number.

Пошук за реєстраційним номером CAS

Сполуки шукаються за реєстраційним номером CAS.

Пошук за енергетичними характеристиками іонів

При пошуку за енергетичними характеристиками іонів (Ion Energetics Searches) запит може ґрунтуватися на значенні енергії, діапазоні значень чи молекулярній формулі іона. Цифрові значення задають в електрон-вольтах (eV). У разі одиничного значення точність пошуку становить 0,05 еВ. У діапазоні значень початкове й кінцеве значення подаються через кому. Треба зазначити, що до результатів пошуку потрапляють лише перевірені дані, що становить близько 40 % доступного обсягу даних.

У цьому розділі доступні такі види пошуку: за енергією іонізації; за спорідненістю до електрона; за спорідненістю до протона; за кислотністю; за енергією відриву електрона.

Пошук у коливальних та електронних спектрах

Цей пошук (Vibrational and Electronic Spectra Searches) можливий у двох варіантах – за значенням коливальної енергії

та на основі електронних рівнів енергії. До результатів пошуку не входять дані за ІЧ-спектрами у газовій фазі.

Дані для пошуку можуть бути зазначені у вигляді одиничного значення або діапазону значень у зворотних сантиметрах ($см^{-1}$). У діапазоні значень початкове й кінцеве з них подаються через кому; точність пошуку становить 5 $см^{-1}$.

Пошук за структурою

Пошук за фрагментами структури молекули (Structure Based Searches) є унікальною властивістю цього сервісу. Такі можливості в основному доступні у платних системах доступу до баз даних.

Процес пошуку тоді займає більше часу, оскільки вимагає значно потужніших комп'ютерних ресурсів. Пошук обмежений тільки тими сполуками, які мають у базі даних структурну інформацію. Тут можливі три види пошуку: заснований на віртуальному хімічному редакторі, заснований на готових файлах і пошук на основі структурних класів.

Пошук за допомогою віртуального хімічного редактора

Цей вид пошуку (Applet Based Structure Search) використовує Java-аплети, які завантажуються у браузер і дають змогу користувачеві побудувати фрагмент структури, котрий він хоче знайти. У результати потрапляють ті сполуки, котрі у своїй структурі мають вказаний фрагмент.

Пошук за допомогою файлів, що містять геометричну інформацію

Цей вид пошуку (File Based Structure Search) схожий до попереднього за винятком того, що на сервер завантажується файл, який містить структурну інформацію. Файл повинен мати MDL mol-формат.

Пошук за структурним класом

Цей вид пошуку (Structure Class Search) дає змогу знайти структури, ґрунтуючись на певних їх характеристиках, що обираються на сторінці пошуку. Доступні такі опції:

– кількість атомів у структурі (за винятком гідрогену);

– кількість зв'язків у структурі;

– кількість циклів у структурі;
– хімічна формула;
– кількість певних функціональних груп у структурі.

У кожній структурі певні атоми можуть належати до різних функціональних груп. Наприклад, карбонова кислота буде мати функціональні групи OH і C=O.

Пошук за молекулярною масою

(Molecular Weight Search). Використовуючи цей вид пошуку, можна задавати конкретне значення молекулярної маси або діапазон значень, розділених комою. Точність пошуку становить 0,5 атомних одиниць маси. Для розрахунку маси використовують атомні маси найбільш поширених ізотопів, хоча на інформаційних сторінках вказується молекулярна маса, що ґрунтується на середніх значеннях атомних мас. Тому результати пошуку та одержані результати можуть дещо відрізнятися.

Пошук реакцій

Цей вид пошуку (Reaction Search) дає змогу шукати інформацію безпосередньо в базі даних хімічних реакцій. Реагенти можна вказувати у вигляді хімічної формули, хімічної назви, реєстраційного номера CAS.

Серед додаткових умов пошуку – можливість вибирати напрям реакції, відзначати додаткові реагенти і продукти, не вказані в запиті, усувати з запиту іони або вибрати специфічний тип реакцій.

Авторський пошук

Авторський пошук (Author Search) дає змогу шукати авторів, на котрих є посилання в базі даних. У формі пошуку можна задати до трьох різних авторів, а також рік публікації або діапазон років через кому. Спочатку вказується прізвище, а потім ініціали, наприклад Smith, D. A. В результаті пошуку можуть бути знайдені автори, праці котрих ще не поміщені в базу даних.

Непрямий пошук

Під непрямим пошуком (Indirect Searches) маємо на увазі навігацію за гіперпосиланнями. Елементи результатів пошуку,

на які в базі даних є додаткова інформація, забезпечені гіперпосиланнями на цю інформацію.

Представлення даних

Більшість даних на цьому сайті доступні в табличній формі. Коли можливо, дані представити в графічній формі, додається потрібне гіперпосилиння.

Таблиці даних

Таблиці, у яких представлені дані, можуть бути різними за конфігурацією, але у них є і спільні риси. Якщо можливо, у таблиці подають гіперпосилання на використані методи та на літературні джерела, при потребі – також коментарі.

Графічне представлення даних

Деякі типи даних можуть бути представлені у графічній формі. Для цього потрібно, щоб браузер підтримував Java-аплети й проксі-сервер, не блокував обмін даними між Java-програмою та сервером NIST.

Якщо Java не працює, система може представляти ІЧ, УФ і мас-спектри у вигляді растрових зображень.

При активному Java спектрами можна оперувати – змінювати масштаб, діапазон значень та ін.

Інформація про рідини

Фазові діаграми рідин можуть бути представлені у графічній формі. При цьому кожна ділянка позначається різними кольорами: синій – рідина, червоний – газ, зелений – надкритичний стан, фіолетовий (magenta) – крива насичення.

Графіки функцій

Деякі дані можуть бути апроксимованими певними функціями, які описують їхню поведінку в певних межах. У такому разі можна побудувати графіки цих функцій та одержати оцінку їх параметрів. Для побудови графіків можна вибирати різні шкали: логарифмічну, напівлогарифмічну або нормальну.

Доступна інформація

Для кожної сполуки можуть бути доступними кілька типів даних, при видачі результатів вони відділяються горизонтальними лініями. Загальна інформація про сполуку подається

зазвичай на початку сторінки, а коментарі та посилання – у кінці.

Загальна інформація

Для всіх сполук існує певний набір даних, відображених на початку сторінки. Ця інформація містить: назву сполуки, молекулярну масу, реєстраційний номер CAS, зображення структури молекули, альтернативні назви, гіперпосилання на іншу інформацію на сторінці й інших сторінках, опції для вибору системи вимірювань.

Посилання на літературу

Посилання на цитовану літературу розміщено в кінці веб-сторінки в порядку цитування в тексті сторінки. Посилання на табличні дані виділені жирним шрифтом.

ПОШУКОВІ МАШИНИ ПАТЕНТНОЇ ІНФОРМАЦІЇ

Як уже зазначалося вище, значну частку первинної наукової інформації складають патенти. Припускають, що більше 70 % науково-технічної інформації, розміщеної в патентних документах, є унікальною і не публікується в будь-яких інших джерелах. Більшість патентних баз даних містить описи винаходів у формі рефератів, формул винаходів або даних Міжнародної патентної класифікації. Деякі бази даних містять також малюнки, схеми й повні тексти патентів. Перераховані факти зумовлюють використання патентних баз даних як джерел інформації, потрібної, наприклад, для моніторингу технічного рівня розробок і винахідницької активності компаній-конкурентів.

У галузі технічних інновацій інформація про запатентовані об’єкти украй потрібна для розробки потенційно нових продуктів. Як засіб моніторингу інновацій, оцінки рівня техніки і технічних рішень, аналізу міжнародних ринків і діяльності компаній-конкурентів, патентна інформація є найоб’єктивнішим джерелом. Патентна інформація сприяє не тільки захисту винаходів, а й, охоплюючи технічні, правові й комерційні аспекти, потрібна в процесі планування бізнесу.

Відомості, які розміщені у патентних базах даних

26.1.1.1. Правові аспекти

До інформації про правові аспекти належать, наприклад, дати подання заявок, дати отримання патентів, прізвища винахідників, назви фірм-патентовласників, пріоритетні дані, відомості про патентні сімейства, відомості про правовий статус та ін. Ця інформація вказує, чи винахід вже запатентований, чи на нього тільки подана заявка. Використання патентної інформації допоможе уникнути дублювання вже виконаних розробок і досліджень, заразом заощадити час і гроші. Моніторинг патентної інформації забезпечить також контроль за можливими порушеннями у використанні Ваших власних патентів.

В Інтернеті наявна низка сервісів, що забезпечують доступ до патентної інформації. Ми розглянемо такі: STN International, американську USPTO та європейську esp@cenet.

FIZ Karlsruhe

Одним із найвідоміших сервісів є спеціалізований інформаційний центр FIZ Karlsruhe, який через свою інформаційну мережу STN International надає доступ до найбільших міжнародних і національних патентних баз даних. У цілому в STN представлено 26 баз даних, що містять відомості про патентні документи і товарні знаки.

Міжнародна мережа наукової і технічної інформації STN International є онлайновою службою, що надає доступ до більш ніж 210 баз даних зі всіх галузей науки, техніки і промисловості. Сумарний обсяг баз даних перевищує 350 млн документів. Близько 30 млн патентних документів, розміщених у найбільших міжнародних і національних патентних базах даних, доступні в мережі STN, який перебуває в сумісному управлінні FIZ Karlsruhe (ФРН), Реферативної служби з хімії (Chemical Abstracts Service (CAS), Колумбус, США) і Японській науково-технічній корпорації (Japan Science and Technology Corporation (JST), Токіо, Японія).

На жаль, ця база даних є платною і приймає підписку від організацій. Волинський національний університет імені Лесі

Українки не має доступу до цієї бази даних.

USPTO

Служба патентних і торгових марок США (англ. US Patent and Trademark Office – USPTO, http://uspto.gov) утримує базу даних патентів, до якої надає вільний доступ.

У базі даних доступні повні тексти патентів США, починаючи з 1976 р., та бібліографічні описи патентів упродовж 1790–1975 рр. Для патентів до 1976 р. достовірність текстів не гарантується, але вони подані. Можна припустити, що відбувається поповнення бази даних, і нижня межа повністю доступних документів буде знижена. Для патентів зазвичай доступні копії сторінок патентів у форматі TIFF (для перегляду потрібно встановити доповнення до браузера, на сторінці перегляду зображень є відповідні посилання).

База даних патентів оновлюється кожен вівторок за винятком державних свят США. Класифікаційні індекси оновлюють кожні два місяці.

Кожен патент має власну веб-адресу.

Приклад 57. *Адреса патенту в USPTO:*

http://patft1.uspto.gov/netacgi/nph-Parser?patentnumber=4185078, за цією адресою відкриється сторінка, присвячена патенту № 4185078.

Пошук патентів

Варто зазначити, що на головній сторінці USPTO немає ні форми для пошуку патентів, ні безпосередніх посилань. До неї можна перейти за допомогою навігації: у розділі “Patents” є підрозділ “Search”, на сторінці, що відкрилася (http://www.uspto.gov/patents/process/search/index.jsp), вибрати потрібний пункт або безпосередньо ввести в адресний рядок браузера сторінку вибору методів пошуку (http://patft.uspto.gov/), до якої немає безпосереднього гіперпосилання з головної сторінки. На цій сторінці містяться посилання на форми простого пошуку (Quick Search), розширеного пошуку (Advanced Search) та пошуку за номером патенту (Patent Number Search) для виданих патентів (Issued

Patents, зліва) і використання патентів (Patent Applications, справа).

Розглянемо особливості цих пошукових форм більш детально.

Швидкий пошук

Сторінка швидкого пошуку (Quick Search Page) містить пошукову форму з двох полів, у яких вводяться пошукові терміни. Друге поле заповнювати не обов'язково. Для кожного терміна можна задати одну з 30 ділянок пошуку (або за всіма полями). Терміни можуть бути пов'язані логічними операторами AND, OR або ANDNOT. Меню, розміщене знизу, дає змогу вибрати діапазон років для пошуку: з 1976 р. (повнотекстовий пошук) і з 1790 р. (вся база даних).

Пошук не бере до уваги регістр букв. Як термін можна задавати фразу, але її тоді треба подати в лапках. Пошукова машина не проводить морфологічного аналізу введених слів – у якій формі вони введені, такі вона й буде шукати. Але є можливість пошуку слів, які однаково починаються – для позначення будь-якого закінчення слова використовують знак долара "$" (початок слова повинен містити не менше чотирьох літер, при пошуку в окремих полях – не менше трьох).

Сторінка результатів пошуку містить близько 50 знайдених патентів – номер і назва патенту. Патенти у списку результатів згруповані за новизною – спочатку найновіші, потім – старіші.

Список знайдених патентів можна переглядати по 50 найменувань на сторінці, натискаючи відповідні кнопки, або вказати номер, з якого будуть показані результати пошуку.

Розширений пошук

Сторінка розширеного пошуку слугує для більш детального задання умов пошуку. Сторінка містить опцію вибору діапазону років та єдине поле для вказування запиту, котрий формується з використанням термінів, службових команд і дужок. Максимальний обсяг запиту може становити близько 256 символів. Ознакою службової команди слугує символ слешу "/". У нижній частині міститься список службових команд для уточнення пошуку.

Сторінка результатів пошуку така, як і при звичайному пошуку: вона містить поле для уточнення пошуку, запит, який

потрібно подавати в тому ж форматі, що й на формі пошуку. Єдина відмінність – пошук буде проводитися у тому ж діапазоні дат, що був заданий спочатку.

Розробники попереджають, що база даних не призначена для масового пошуку та завантаження патентів (більше 10 000 запитів за день або перегляд більше ніж 1000 патентів) – у цьому разі користувачі будуть блокуватися без попередження.

Отже, USPTO є зручним засобом для пошуку в патентній базі США, котра є світовим лідером у створенні наукової продукції.

esp@cenet

Патентна служба esp@cenet (англ. Europe's Network Patent Databases – мережева база даних патентів Європи, http://ep.espacenet.com) надає вільний доступ до фондів Європейського патентного бюро EPO (англ. European Patent Office). Вона має доступ до інформації про більш ніж 65 млн патентів з 92 країн світу (починаючи з 1836 р), більше 18 млн з яких мають реферат англійською мовою. Заявлений обсяг бази даних – 350 млн сторінок формату А-4. Робочі мови сайту – англійська, французька, німецька, японська. Він працює цілодобово, проте з понеділка по суботу з 5^{00} до 5^{15} год за середньоєвропейським часом ефективне використання його не гарантовано. До баз даних сайту доступ пошукових роботів обмежений.

Із головної сторінки доступні чотири різні форми пошуку: простий (Quick Search), розширений (Advanced Search), пошук за номером (Number Search) та пошук в класифікаційній системі (Classification Search). У всіх чотирьох формах потрібно вибрати у меню одну базу даних патентів із трьох: EP (європейська колекція), Worldwide (повна колекція опублікованих патентів із 92 країн світу) або WIPO (патенти Всесвітньої організації інтелектуальної власності – World Intellectual Property Organization).

Під час пошуку можна використовувати додаткові знаки:

– зірочку "*" – для пошуку рядка символів будь-якого обсягу;

– знак питання "?" – для позначення однієї або більше літер;

– решітку “#” – для позначення тільки однієї літери.

Проте у використанні додаткових знаків є певні обмеження – вони не можуть використовуватися всередині слова, а лише наприкінці; їх можна використовувати тільки в полях “Title”, “Title or abstract”, “Inventor or Applicant”. Слово, що містить додатковий знак, повинно містити мінімум дві літери, тоді кількість послідовних знаків обмежена трьома. Якщо слово містить три й більше літер на початку, то максимальна кількість знаків “?”, “#” у слові не повинна перевищувати семи (три – для “*”).

Розробники попереджають, що використання в запитах таких знаків значно збільшує час пошуку й може призвести до відмови сервера. Оскільки пошукові форми різних видів пошуку мають свої особливості, розглянемо їх більш детально.

Максимальна кількість пошукових термінів у кожному полі не повинна перевищувати чотирьох слів. Недопустиме використання у пошукових формах апострофів, лапок, слешів, слова мають відділятися пропусками. При потребі можна використати логічні оператори “OR” або “AND”, за замовчуванням використовується оператор “AND”.

Простий пошук

Форма простого пошуку (Quick Search) призначна для пошуку у відповідній базі патентів за ключовим словом, прізвищем автора чи організації. Вона складається з трьох пунктів:

– бази даних (можливий вибір обговорювався вище);

– типи пошуку. Тут треба вказати, що потрібно шукати: слова у заголовках чи рефератах, персони або організації, слова в повному тексті патентів;

– поля для введення ключових слів.

Розширений пошук

Форма розширеного пошуку (Advanced Search) призначена для пошуку у відповідній базі патентів з використанням усіх можливих опцій. Форма пошуку складається з двох пунктів:

– база даних (можливий вибір обговорювався вище);

– пошукових термінів – цей пункт містить декілька полів, у які можна вписувати ключові слова:

1) key word(s) in title (ключові слова в заголовку);

2) key word(s) in title or abstract (ключові слова в рефераті);

3) key word(s) in full text (ключові слову в усьому тексті);

4) publication number (номер публікації);

5) application number (номер використання патенту);

6) priority number (номер пріоритету);

7) publication date (дата публікації);

8) applicant(s) (користувачі патентом);

9) inventor(s) (винахідники);

10) European Classification (ECLA) (європейська класифікація патентів);

11) International Patent Classification (IPC) (міжнародна класифікація патентів).

Терміни, вказані в різних полях, завжди пов'язуються логічним оператором "AND".

Пошук за номером патенту

Такий вид пошуку (Number Search) призначений для пошуку за номером патенту, застосування, пріоритету чи посилання на непатентну літературу (NPL – англ. Non-patent literature) у відповідній базі патентів. Форма пошуку складається з двох пунктів:

– бази даних (можливий вибір обговорювався вище);

– вказаного номера (у цьому полі потрібно ввести номер патенту, публікації чи посилання з використанням коду країни).

Номер публікації складається з двобуквеного коду країни та серійного номера, що містить від однієї до 12 цифр. Номер застосування складається з двобуквеного коду країни, року заповнення (чотири цифри) та серійного номера, який містить від однієї до семи цифр. Номер пріоритету складається з двобуквеного коду країни та серійного номера з семи цифр. Номер посилання складається з коду непатентної літератури та серійного номера. Максимально допустима кількість номерів у цьому полі – чотири.

Пошук класифікації

Пошук класифікації (Classification Search) призначений для пошуку або навігації у європейській класифікації патентів.

Європейська система класифікації патентів (ECLA – англ. The European Classification system) – ієрархічна класифікаційна

система, яка застосовується до опублікованих патентів. Вона містить близько 134 000 різних груп патентів. Основні розділи класифікації (літера та розділ):

A – human necessities (людські потреби).

B – performing operations; transporting (виконання дій, транспортування).

C – chemistry; metallurgy (хімія, металургія).

D – textiles; paper (текстиль, папір).

E – fixed constructions (нерухомі конструкції).

F – Mechanical engineering; lighting; heating; weapons; blasting engines or pumps (механічні розробки, освітлення, нагрівання, зброя, реактивні двигуни чи насоси).

G – physics (фізика).

E – electricity (електрика).

Y – general tagging of new technological developments (загальне маркування та нові технологічні розробки).

Пошукова форма містить поля для введення ключових слів та опції відповідно до основних розділів Європейської класифікації патентів. Тут можна провести пошук за ключовим словом або перейти за гіперпосиланням до відповідного підрозділу класифікації.

Результати пошуку

У разі успішного пошуку можна знайти список результатів, до якого входить така інформація: критерії пошуку, кількість і список знайдених документів. До кожного знайденого патенту вказана назва, що містить гіперпосилання на сторінку документа, опція для внесення результату в особистий список патентів та відповідні бібліографічні дані. Якщо у результатах трапляються ключові слова, вони маркуються певним кольором.

На запит може бути знайдено максимум 100 000 документів, з них може бути показано максимум 500 результатів. На одній сторінці у скороченій формі розміщено 30 результатів, на розгорнутій – 15. Для переключення між розгорнутою і короткою формами слугує пункт “Compact”.

Для кожного знайденого документа подається обмежена кількість інформації, якщо показано не все, число в дужках, наприклад (+3), показує кількість елементів інформації, що не

наведені, але будуть доступні при переході на сторінку відповідного патенту.

Список результатів можна згрупувати за датою завантаження, датою пріоритету, ім'ям винахідника, ім'ям власника патенту чи за класифікацією "ECLA", якщо кількість результатів не перевищує 500.

Знизу та зверху сторінки результатів пошуку містяться опції управління списком результатів: подати список у компактній формі, роздрукувати, експортувати чи уточнити пошук. Результати можуть бути експортовані в CSV-файл, котрий можна переглянути за допомогою програми MS Excel.

Сторінка документа

Сторінка опису патенту містить низку закладок: бібліографічні дані, опис патенту, формулу винаходу, зменшені зображення малюнків, зображення оригінального документа, статус INPADOC.

26.4.6.1. Бібліографічні дані

Бібліографічні дані (Bibliographic data) документа складаються із заголовка та реферату, а також містять номер публікації (з кодом класифікації в дужках), дату публікації, стандартизовані імена винахідника і правонаступника (зі знаком "+" доступні альтернативні записи імен), інтернаціональна та європейська класифікації патенту, номер застосування, номер пріоритету а також гіперпосилання до споріднених патентів за INPADOC і списку цитованих документів.

У пункті "Also published as" подано альтернативні джерела наведеної в патенті інформації.

У тексті може існувати кнопка "Translate this text", котра приводить до сторінки, де можна вибрати мову перекладу. Тест може бути перекладений на вибір німецькою, французькою, італійською або іспанською мовами. Якщо текст не англійський, то він може бути перекладений також англійською мовою. Ця кнопка може міститися на будь-якій закладці, де є текст.

26.4.6.2. Опис патенту

Опис патенту (Description) містить такі компоненти:

вказівку на галузь техніки, до якої належить відкриття; невеликий літогляд для розуміння актуальності відкриття; зміст відкриття (описуються технічні проблеми, методи їх вирішення та переваги запропонованого відкриття); короткий опис графіків і малюнків; детальний опис принаймні одного методу реалізації винаходу; пропозиції щодо використання в промисловості.

26.4.6.3. Формула винаходу

Формула винаходу (англ. Claim – заява, або претензія на винахід) є частиною патенту, що визначає межі законного захисту, заявленого для винаходу. Для їх інтерпретації використовують опис і малюнки.

У тексті є кнопка “Claims tree”, котра дає змогу представити пункти формули у вигляді ієрархічного дерева з можливістю навігації за рівнями.

26.4.6.4. Мозаїка

Мозаїка (Mosaics) – зібрання зменшених копій сторінок з ілюстраціями документа. Оскільки ці зображення менші за розміром, ніж оригінальна сторінка, вони простіші для сприймання користувачем. На одній сторінці може міститися близько шести зображень.

26.4.6.5. Оригінальний документ

На цій закладці (Original documents) можна переглянути зображення сторінок оригінальних документів, роздрукувати, а також завантажити повністю документ у PDF-форматі. Кількість сторінок патентного документа, доступного для завантаження не повинна перевищувати 250.

26.4.6.6. Описи кодів

Закладка кодів бази даних (INPADOC legal status) містить детальні пояснення й описи кодів легального статусу INPADOC. Нині доступні коди країн Євросоюзу, Великобританії та Німеччини, ведеться робота над залученням Австралії, а потім й інших країн.

Веб-сервіс esp@cenet є унікальним інструментом для пошуку та перегляду патентів зі всього світу. Його обов'язково слід використовувати при підготовці літогляду до курсової, випускної, дипломної чи магістерської роботи.

РОБОТА З ДЖЕРЕЛАМИ ІНФОРМАЦІЇ

Пошук наукової інформації зазвичай не є самоціллю: знайдена інформація використовується в науковій, методичній і педагогічній роботі, під час написання літературного огляду чи реферату. Джерел інформації накопичується багато і постає питання ефективного їх використання.

Працюючи зі знайденими джерелами інформації, бажано дотримуватися таких правил:

1. Слід читати тільки джерела інформації, що безпосередньо стосується теми. Все другорядне можна переглянути. Ознайомлюючись з наступним джерелом, треба постійно ставити собі питання: "Чи дійсно мені потрібне те, що я зараз читаю?". Завжди виникає спокуса звернути увагу на щось цікаве, хоч і не по темі. Як засвідчує практика, кінцева тематика читаного завжди майже однакова. Підраховано, що для переходу з будь-якої сторінки в Інтернеті до іншої потрібно використати не більше 19 переходів за гіперпосиланнями.

Перед тим, як ознайомитися з інформацією, бажано проглянути її джерело. Для активізації розумової діяльності при читанні слід ставити, а краще записувати, питання, відповіді на які треба знайти під час читання.

Зміст будь-якого не до кінця зрозумілого слова варто уточнити в тлумачному словнику або в енциклопедії. Активізація словникового запасу згодом допоможе під час написання роботи.

Для підвищення ефективності засвоєння інформації потрібно більше читати – тільки постійне тренування може гарантувати підвищення швидкості читання і стійке інтелектуальне зростання.

Для хіміка знання англійської мови обов'язкові, оскільки майже вся наукова інформація з хімії в світі представлена англійською мовою. Якщо у школі чи в університеті не

вивчалася англійська мова, слід почати вивчати її самому. Насамперед потрібно засвоїти основні граматичні конструкції, потім – терміни, далі – інші слова.

Типова помилка студентів, коли вони натрапляють на англомовне джерело – повнотекстовий переклад. Звичайно, переклад англійських текстів – хороша практика в оволодінні цією мовою, але тільки тоді, коли вона проводиться під контролем спеціаліста-педагога, ознайомленого зі специфікою англомовної літератури з хімії. Студент перекладає текст за допомогою комп'ютера іншого студента-лінгвіста. При цьому часто губиться сенс інформації. Оскільки структура статей різними мовами однакова, то є можливість вибрати зі статті потрібну частину інформації, а тоді розпочати його безпосередній переклад більш уважно. Маючи певну практику, іноді можливо зрозуміти сенс інформації навіть без перекладу. Потрібно пам'ятати, що важливо не перекласти інформацію з джерела, а зрозуміти її сенс чи вибрати потрібну її частину.

Написання конспектів, опрацювання матеріалів книг і періодики

Вивчення літературних джерел передбачає насамперед їх читання, якому зазвичай передує перегляд – ознайомлення з анотацією, змістом, передмовою. Читання може бути ознайомлювальним і поглибленим. Під час ознайомлювального читання слід з'ясувати, які відомості, потрібні для дослідження, містяться в документі. Цей етап роботи зазвичай відбувається на спочатку, коли потрібно відібрати літературу з теми, скласти план роботи. Попередній перегляд і ознайомлювальне читання доречно супроводжувати складанням плану документа і виписок з нього. Метою поглибленого читання є з'ясування основних положень роботи, системи доказів, висновків автора.

Працюючи з літературними джерелами, рекомендується вести записи прочитаного. Записи поділяють на систематизовані й несистематизовані. Існують такі прийоми ведення систематизованих записів: плани, конспекти, реферати, анотації. До несистематизованих записів належать виписки.

План – це систематизований запис тексту, що відображає структуру документа.

План складається при ознайомлювальному читанні, тоді, коли ще точно не відомо, які розділи джерела інформації стануть в нагоді надалі. План допоможе пригадати сенс інформації й визначити доцільність її подальшого вивчення.

Конспект – це систематизований запис тексту, що відображає зміст документа і його структуру.

Конспект – це стислий виклад доповіді, статті, книги. Його застосовують для фіксації сприйнятого матеріалу і для полегшення його подальшого опрацювання. Рекомендується складати конспект за такою схемою: виписувати тільки те, що справді потрібно; у конспекті можна використовувати різні скорочення й умовні позначення, як загальноприйняті, так і власні; ефективним також буде використання логічних схем, особливо особам з абстрактним мисленням.

Слід завести для конспектів окремі зошити, блокноти або картки. На обкладинці записується номер зошита, на звороті обкладинки – джерела інформації, конспекти яких містяться в зошиті, за порядком їх опрацювання. Така організація конспектування в подальшому спрощує пошук і систематизацію матеріалу.

У конспектах зазвичай залишають широкі поля, щоб згодом робити на них помітки. Використовують також різні виділення важливих думок маркером, підкреслення кольоровими ручками.

Під час оформлення конспектів важливим є внесення у них такої інформації:

1. Бібліографічний опис конспектованого джерела (бажано в повній формі). Тоді не доведеться під час написання літогляду знову шукати джерела інформації тільки для того, щоб виписати описи.
2. Коротка анотація джерела інформації, котра резюмує її зміст. Якщо виникне потреба використати конспект через декілька років, без цього буде дуже складно пригадати зміст статті за конспектом.

Можна конспектувати і в комп’ютері. В цьому разі слід приділити увагу збереженню резервних копій на випадок втрати інформації. Найбільш плідним тут є використання

спеціалізованих бібліографічних баз даних. Звичайно, можливо, спочатку буде важко навчитися з ними працювати, проте в майбутньому використання цих навиків значно спростить роботу. З більш звичних програм можна використовувати MS Word, MS Excel, MS Access, MS Binder, хоча робота з ними, особливо при накопиченні великого обсягу інформації, буде не дуже зручною. Крім того, конспектуючи у комп'ютері, бажано сформувати для себе картотеку бібліографічної інформації, роздруковуючи конспекти на папері. У разі втрати інформації у комп'ютері буде можливість відновити її, або використати паперовий варіант, тим більше, що під час написання роботи часто використовується далеко не вся знайдена і законспектована інформація.

При підготовці конспект є найбільш поширеною формою систематизованих записів прочитаного, він допомагає накопичити потрібний матеріал. Конспект може бути коротким або розгорненим. У короткому конспекті фіксують основні тези роботи, в розгорнутому крім тез подається докладний виклад роботи або її окремих частин.

Складаючи конспект, рекомендується стисло, без зайвих слів записувати найважливіше з прочитаного твору. В основі конспекту повинен лежати план твору, що вивчається. Тому корисно насамперед скласти план, а потім, розкриваючи його пункти, вести записи. Конспектують документи зазвичай своїми словами. Найбільш важливі і найбільш важкі для розуміння частини тексту краще цитувати. Обов'язково потрібно цитувати методики, причому методики іноземними мовами бажано не перекладати, а виписувати дослівно. Під час перекладу може втратитися важлива інформація, тому завжди варто мати оригінал. Конспект рекомендується вести так, щоб легко можна було знайти потрібний матеріал. Писати треба на одній стороні сторінки, обов'язково відзначаючи в конспекті номери сторінок джерела, з яких узято відомості.

Реферат – максимально стислий виклад змісту документа або його частини. Він повинен містити основні ідеї, фактичні відомості та висновки документа.

Реферат за своїм змістом близький до твору на тему, але у творі висловлено власні міркування про предмет викладу, а в рефераті переказують матеріали джерел.

Виписки – це короткі записи окремих фрагментів текстів, цифр, дат й інших фактів.

Для виписок краще завести блокнот або зошит.

Ф. А. Кузін [19] пропонує такі етапи роботи над літературними джерелами:

– загальне ознайомлення з джерелом в цілому за його змістом;

– поверхневий перегляд всього вмісту документа;

– послідовне читання матеріалу;

– вибіркове читання певної частини твору;

– виписка матеріалів, що Вас цікавлять;

– критична оцінка записаного, його редагування і "чистовий" запис як фрагмент майбутньої власної роботи.

Вивчаючи літературу на певну тему, використовують не усю інформацію, а тільки ту, яка безпосередньо стосується до виконуваної роботи. Тому критерієм оцінки прочитаного є можливість його подальшого практичного використання у своїй роботі.

Осмислення і систематизація матеріалу

Опрацьовуючи літературу, потрібно не тільки копіювати матеріал, а й заразом обдумувати знайдену інформацію. Цей процес може здійснюватися протягом усієї роботи над власною темою – тоді власні думки, що виникли при ознайомленні з чужими роботами, можуть стати основою для отримання нових знань.

Коли накопичено достатньо матеріалу, можна розпочинати його осмислення і систематизацію. Якщо записи зроблено на картках, то їх класифікують за подібністю інформації, прагнучи відшукати змістовні й логічні зв'язки.

Осмислення – це пошук відповідей на питання, що цікавлять.

Відповідно до поставлених завдань під час осмислення матеріалу відшукують відповіді на питання: "що? де? коли?

чому? навіщо?". Відповіді мають бути обґрунтовані, несуперечливі, за винятком випадків ірраціонально-складних феноменів, де можливі ситуації додаткової й діалектичної суперечності. Нерідко осмислення розуміють як опис і пояснення фактів.

Під час роботи над літературними джерелами важливим є їх відбір та оцінка. Бажано використовувати тільки перевірену наукову інформацію.

Дуже корисним прийомом роботи з літературними джерелами є складання картотеки наукових джерел із теми. Грамотно складена картотека навіть при побіжному огляді заголовків дає змогу охопити проблему в цілому.

Бібліографію слід вести і зберігати так, щоб відомості про потрібний документ легко можна було знайти навіть після багатьох років роботи. Бібліографічні картки об'єднують у бібліографічну картотеку за допомогою пристосувань для їх впорядковування, наприклад, клясера, ящика або коробки. Використання комп'ютера змінює лише форму ведення бібліографії, але не принцип. У комп'ютерних програмах також використовують принцип картотеки: бібліографічні відомості заносять в "електронні картки", розміщують їх за алфавітом або групують. Перевагою картотечної форми ведення бібліографії є можливість порівняно легко змінити порядок розміщення, швидко додати нові дані або, навпаки, видалити зайві.

Слід сказати, що ще не уся наукова інформація доступна в електронній формі, тому потрібно користуватися бібліотеками. У бібліотеках сьогодні широко поширений сервіс ксерокопіювання потрібної інформації. Тут треба застерегти від бездумного копіювання знайденої інформації – краще потратити деякий час на те, щоб виокремити дійсно потрібну інформацію.

Ксерокопії знайдених матеріалів слід скріплювати відповідно до певного, конкретного факту та обов'язково записувати бібліографічні дані джерела.

Як читати

У процесі читання літератури обов'язково трапляються нові джерела серед посилань і прикнижних списків використаних робіт, тому потрібна постійна систематизація матеріалу, його

впорядкування відповідно до поставленого завдання. Це можна здійснити, наприклад, за допомогою картотеки, що складається з карток і роздільників. Найкраще сформувати три розділи: “Прочитати”, “Виписки” і “Прочитано”. Створення такої картотеки дає змогу по суті закласти основи майбутніх наукових публікацій. Проте інформація, що міститься у відібраній для вивчення літературі, іноді перевищує дійсні потреби для певної роботи, тому слід заздалегідь виявляти усе потрібне і відкидати зайве. Закладаються елементи вибіркового читання (спочатку поверхневий перегляд джерела, ознайомлення з назвою його розділів і лише потім докладне вивчення вибраного змісту). Важливе значення для роботи з науковою літературою належить організації робочого місця. Робоче місце й інструмент, яким людина працює, мають бути звичні, це скорочує до мінімуму період адаптації. На робочому місці не повинні з’являтися нові предмети (об’єкти), які привертають увагу до себе і відволікають від роботи. Бажано до початку роботи продумати й оцінити, що може бути потрібно у процесі роботи, щоб потім не шукати для себе приводу перервати розпочату справу. При роботі з літературними джерелами потрібно уміти правильно читати, розуміти і запам’ятовувати прочитане. Учені виокремили чотири основні способи обробки інформації при читанні: побуквенне, поскладове, словесне (є видимим перший склад першого слова і перші букви другого слова, решта частини слова вгадується), за поняттями (з тексту вибирають тільки окремі ключові слова, а потім синтезують думку, що міститься в одній або декількох позиціях). Читання за поняттями характерне для людей, що мають певні навики, великий обсяг знань для розуміння матеріалу й хорошу пам’ять. Для розуміння складного тексту слід не тільки бути уважним при читанні, мати знання та уміти їх застосовувати, а й володіти певними розумовими прийомами. Один із них полягає в можливості сприймати не окремі слова, а речення і навіть цілі групи речень, тобто абзаци. При цьому використовують так звану антикипацію – смислову здогадку. Людина, що швидко читає, зазвичай за декількома буквами вгадує слово, за декількома словами – фразу, за фразами – сенс усього абзацу. Потрібно навчитися читати саме так матеріал, що вивчається. Для цього можна

використовувати так званий диференціальний алгоритм, відповідно до якого обробка кожного абзацу починається з виявлення ключових слів, що несуть основне смислове навантаження, після чого будуються смислові ряди, тобто відбувається стиснення тексту способом виділення ключових слів і формування на їх основі лаконічних виразів (своєрідне "просіювання" тексту, у результаті якого можна його зрозуміти). Після стиснення тексту відбувається процес його якісного перетворення – і в оброблюваному повідомленні виявляється тільки дійсне значення його змісту. Навчаючись швидкого читання, ставиться завдання формування нових звичок, які прискорюють читання. Перший спосіб підвищення швидкості читання – виявлення прихованих резервів мозку, активізація процесів мислення при читанні. Один із методів вирішення цього завдання – використання алгоритмів. Інтегральний алгоритм визначає послідовність дій при сприйнятті тексту. Слово "інтегральний" означає, що дія алгоритму розповсюджується на весь текст. Для використання інтегрального алгоритму потрібно запам'ятати усі його блоки, розуміти та уявляти собі їх зміст. Сучасна структурна лінгвістика стверджує, що усі суспільно-політичні й науково-технічні тексти володіють надмірною надлишковістю – близько 75 %. Знайти і зосередити увагу при читанні на змістовних елементах тексту допомагає інтегральний алгоритм читання. Під час його використання значно скорочується час обробки неінформативних елементів тексту.

ЛІТЕРАТУРА

1. ABC Chemistry: Азбука Web-поиска для химиков / А. А. Рагойша [Электронный ресурс] – Режим доступа : http://www.abc.chemistry.bsu.by/default.htm

2. Boell S. K. A Scientometric Method to Analyze Scientific Journals as Exemplified by the Area of Information Science., 2007 Master Thesis thesis, Saarbrücken : Saarland University. [Thesis] / S. K. Boell. [Electronik resourse]. – Mode of access : http://eprints.rclis.org/15610/3/Boell%2C_Sebastian_K-2007-Master_Thesis.pdf. – [Наукометричний опис законів розподілу кількості вчених].

3. Bradford S. C. Documentation / S. C. Bradford. 2-nd ed. – London : Lockwood, 1953. – P. 154. – [Розсіювання інформації].

4. Burton R. E. The “half-life” of some scientific and technical literatures // American Documentation. – 1960. – Vol. 1. – P. 98–109. – [концепція старіння наукової інформації].

5. Deja Vu: a Database of Highly Similar Citations: A study of scientific publication ethics [Elektronik resourse]. – Mode of access : http://spore.vbi.vt.edu/dejavu/. – 30.03.2010. – [Пошук подібних статей в базі даних PubMed].

6. Karen L. M. Origin-The Scholarly Journal / L. M. Karen [Elektronik resourse]. – Mode of access : http://www.slais.ubc.ca/COURSES/libr500/fall1999/WWW_presentations/K_MacDonell/origin.htm. – [Про перші наукові журнали в світі].

7. Nikos Drakos. The explosion of scientific journals revised and updated by: Marcus Hennecke, Ross Moore, Herb Swan, Jens Lippmann, Marek Rouchal, Martin Wilck and others / Nikos Drakos [Elektronik resourse]. – Mode of access : http://www.science.uva.nl/projects/commphys/papers/thesisfh/chapters/node4.html CBLU, University of Leeds, March 2nd, 1998.

8. Regional totals for R&D Expenditure (GERD) and Researchers, 2002 and 2007 / UNESCO Institute for statistic: Data Centre [Elektronik resourse]. – Mode of access : http://stats.uis.unesco.org/unesco/ReportFolders/ReportFolders.aspx. – 31.03.2010. –

[Дані ЮНЕСКО про вчених у світі].

9. Search Engine Market Share [Elektronik resourse]. – Mode of access : http://marketshare.hitslink.com/search-engine-market-share.aspx?qprid=4. – [Рейтинг пошукових серверів світу].

10. Влияние солнечной активности на работу ученых / Л. Завальский : блог [Электронный ресурс]. – Режим доступа : http://www.inauka.ru/blogs/article40670. – 27.05. 10.

11. Гречихин А. А. Библиографическая евристика : учебник / А. А. Гречихин [Электронный ресурс]. – Режим доступа : http://www.hi-edu.ru/e-books /BibliogEvrist/index.htm. – [Види інформаційного пошуку].

12. Гречихин А. А. Библиографическая евристика : учебник / А. А. Гречихин [Электронный ресурс]. – Режим доступа : http://www.hi-edu.ru/e-books/BibliogEvrist/index.htm

13. ДСТУ 7.11-2004 Система стандартов по информации, библиотечному и издательскому делу. Библиографическая запись. Сокращение слов и словосочетаний на иностранных европейских языках [Текст] : (ГОСТ 7.11-2004); введен 2005-09-01. – М. : Изд-во стандартов.

14. ДСТУ 7093:2009 Система стандартів з інформації, бібліотечної та видавничої справи. Бібліографічний запис. Скорочення слів і словосполук, поданих іноземними європейськими мовами [Текст] : (ГОСТ 7.11-2004); надано чинності 2010-04-01. – К. : Держспоживстандарт України, 2009. – 89 с.

15. ДСТУ 7093-2009. Система стандартів з інформації, бібліотечної та видавничої справи. Бібліографічний запис скорочення слів і словосполук, поданих іноземними європейськими мовами [Текст]. – На заміну ГОСТ 7.11-78 ; надано чинності. 2010-04-01. – [Б. м. : б. в.]. – 90 с. – Б. ц.

16. ДСТУ ГОСТ 7.1:2006 Бібліографічний запис. Бібліографічний опис. Загальні вимоги та правила складання [Текст] : (ГОСТ 7.1-2003); введент 2007-07-01. – К. : Держспоживстандарт України, 2009.

17. Дягилева Т. Д. Книжная культура в информационную эру / Т. Д. Дягилева, Н. А. Зинкевич НБ СПбГУП // Библиотечные фонды: проблемы и решения : Электронный

журнал-препринт, – 2002. – июнь. – № 3. [Электронный ресурс]. – Режим доступа : http://www.rba.ru/or/comitet/12/mag3/diaglich.html. – [Ріст кількості книг в світі].

18. Капица П. Л. Эксперимент, теория, практика / П. Л. Капица. – М. : Наука, 1974. – 288 с.

19. Кузин Ф. А. Кандидатская диссертация: методика написания, правила оформления, порядок защиты / Ф. А. Кузин. – М. : "Ось-89", 1997. – 208 с.

20. Кузнецов И. Н. Рефераты, контрольные, курсовые и дипломные работы : метод. рекомендации по подготовке и оформлению / И. Н. Кузнецов, Л. В. Лойко. – Мн.: Завигар, 1998. – 146 с.

21. Паршукова Г. Б. Технологии информационного поиска / Г. Б. Паршукова [Электронный ресурс]. – Режим доступа : http://www.spsl.nsc.ru/win/obsemin/infpoisk/Inf_poisk.files/frame.htm. – [Веб-презентація].

22. Потапов В. М. Химическая информация. Где и как искать химику нужные сведения / В. М. Потапов, Э. К. Кочетова.– М. : Химия, 1988. – 224 с.

23. Рагойша А. А. Интернет для начинающих и не только... . – Мн. : Красико-Принт, 2004. – 96 с. – (Книжная серия : "Репетитор").

24. Рагойша А. А. Поиск химической информации в Интернете. Поисковые системы и тематические каталоги : учеб. пособие для студ. хим. фак. / А. А. Райгоша. – Мн. : БГУ, 2003. – Ч. І. – 87 с.

25. Рагойша А. А. Поиск химической информации в Интернете: научные публикации : учеб. пособие для студ. хим. фак. спец. 1-31 05 01 / А. А. Райгоша. – Мн. : БГУ, 2007. – 71 с. – ISBN 978-985-485-575-2.

26. Введення в дію нового стандарту з бібліографічного опису ДСТУ ГОСТ 7.1:2006. Основні відмінності від ГОСТ 7.1.–84 / Устыннькова О., Сенько П., Юлдашева С., Регыдайло Н. [Электронный ресурс]. – Режим доступу : http://www.ukrbook.net/DSTU_pabl.htm. – 25.03.10]. [Особливості застосування нового стандарту].

27. Хавкина Л. Б. Таблица авторских знаков двоичных [Текст] : практ. пособие для библиотекарей / Л. Б. Хавкина ; под ред. Ю. Н. Столярова. – 25-е изд. – М. : ТОО “Либерея”, 1992. – 24 с. – SBN 5-85129-064-1.

28. Харыбина Т. Н. Комплексная методика изучения информационных потребностей пользователей : [опыт Центральной библиотеки Пущинского научного центра РАН] [Текст] / Т. Н. Харыбина, Н. А. Слащева, Ю. В. Мохначева // Научные и технические библиотеки : сб. по вопросам теории и практики библиотечного дела.– 2008. – N 4. – С. 62–71. – ISSN 0130–9765.

29. Чурсин Н. Н. Популярная інформатика / Н. Н. Чурсин. – К., 1982 [Электронный ресурс]. – Режим доступа : http://n-t.ru/ri/ch/pi05.htm. – [Напівперіод старіння інформації].

30. Яблонский А. И. Модели и методы исследования науки / А. И. Яблонский. – М. : Эдиториал УРСС, 2001. – 400 с. – ISBN 5-8360-0257-6.

Навчальне видання

Кормош Ж. О., Супрунович С. В.

Наукова інформація з хімії, її пошук та обробка

Навчальний посібник

Редактор *І. В. Бурко*
Коректори *Л. С. Пащук, В. Є. Сикора, І. В. Бурко*
Технічний редактор *Ю. І. Поліщук*

Підп. до друку 00.00.2011. Формат 60×84$^{1}/_{16}$. Папір офс. Гарн. Таймс. Друк цифровий. Обсяг 14,88 ум. друк. арк., 14,03 обл.-вид. арк. Наклад 000 пр. Зам. 0000. Видавець і виготовлювач − Волинський національний університет ім. Лесі Українки (43025, м. Луцьк, просп. Волі, 13). Свідоцтво Держ. комітету телебачення та радіомовлення України ДК № 3156 від 04.04.2008 р.